An investigation of student writing in Civil Engineering: A corpus linguistics case study in the Middle East

Mathew, Priya

Award date:
2020

Awarding institution:
Coventry University

An investigation of student writing in Civil Engineering: A corpus linguistics case study in the Middle East

By

Priya Mathew

PhD

July 2019

An investigation of student writing in Civil Engineering: A corpus linguistics case study in the Middle East

By

Priya Mathew

July 2019

A thesis submitted in partial fulfilment of the University's requirements for the Degree of Doctor of Philosophy

Certificate of Ethical Approval

Applicant:

Priya Mathew

Project Title:

The Content Teacher Voice: Exploring the Perceptions, Challenges and Possible Solutions to the Academic Writing Problems of Undergraduate Students in an EFL Context

This is to certify that the above named applicant has completed the Coventry University Ethical Approval process and their project has been confirmed and approved as Low Risk

Date of approval:

08 June 2015

Project Reference Number:

P33371

Acknowledgement

First and foremost, I thank the Almighty for giving me the opportunity and the ability to undertake my PhD. His constant love and protection gave me the dedication and perseverance required to complete this doctoral degree.

I am extremely grateful to my Director of Studies, Professor Hilary Nesi, for her continuous encouragement and motivation, and most of all, her faith in my abilities. She has always inspired me to aspire towards the highest academic standards. I acknowledge with gratitude, the support of my Supervisor, Dr Benet Vincent, for his guidance and the endless hours he must have spent to help make my work as meticulous as possible. My supervisory team have done their best to ensure that I regularly interacted with them, both face to face and through innumerable videoconferencing sessions, without which, I could not have completed my thesis. They made sure that I attended enough academic gatherings, including conferences, symposiums and special interest group meetings whenever I came down to the UK.

I am greatly indebted to Middle East College, Sultanate of Oman, for funding my entire PhD programme. I am especially grateful to Mr Lafeer Mohammed, Managing Director, Middle East College (MEC), for believing in my abilities. I acknowledge Dr Kiran. G.R, Dean, MEC, for his tremendous support and inspiration all through these years. I am grateful to Mr Ashwin John, Deputy Dean and Registrar, MEC, for being a tower of support and good will. I must mention Dr Ahmed Nawaz Hakro and Dr Anupam Srivastav, Associate Deans, MEC, for their support and guidance. I am grateful to all my colleagues at MEC who have sometimes surprised me with their immense support and kindness. I place on record the acknowledgement that my PhD dream would not have been fulfilled but for the support of Middle East College and that I consider it both an honour and a privilege to be part of this illustrious institution.

I am extremely grateful to my colleagues at the Centre for Academic Writing, Ms Deepali Bhatnagar and Dr Radhika, S.N. for their ungrudging generosity and steadfast support

during the course of these years. I also thank my colleagues at the Languages Centre for being there for me, especially Dr Tareq Al Damen, my Local Supervisor. I thank my dear college mates who have rooted for me throughout my PhD years and will be so happy for me.

I am grateful to the Doctoral College team at Coventry University for their unwavering support and guidance, especially Dr Rebekah Smith McGloin and Professor Damien Foster. Their frequent visits to MEC were both an inspiration and a constant reminder that perseverance and hard work would see me through. Special thanks to Prof. Sheena Gardner, Dr Simon Smith, faculty members and other PhD students at the Languages Centre, Coventry University.

I now acknowledge my dear family. I acknowledge the wonderful support of my husband and two loving sons, without whose understanding and patience, I could not have made it. My long absences from home, when I had to be in the UK for weeks on end, would not have been easy for them, but I must say they did not complain once and I am so proud of them. I hope my children will also take away a few life lessons from my research journey since they accompanied me through most of it. I am grateful to my siblings who have always wanted their little sis to succeed in her academic endeavours. I have always banked on their loyalty and steadfast support which have pulled me through the rough patches. Special thanks to my mom-in-law for her prayers and enthusiasm.

I dedicate this thesis to my loving parents. Dad and Mom, thank you for this. The credit is all yours.

Priya Mathew

Abstract

An investigation of student writing in Civil Engineering: A corpus linguistics case study in the Middle East

Abstract

This thesis attempts to characterize the two main genres in the Omani Corpus of Academic Writing-Civil Engineering (OCAW-CE) using Corpus Linguistic methods. The design of OCAW-CE allows for the comparison of two different genres within the same discipline, Case Study (CS) and Methodology Recount (MR), making it one of the first studies that have isolated two student academic genres to study their effects on linguistic choices. It addresses three main questions relating to these two genres. The first question is about the move structure of the primary genres in OCAW-CE. A detailed move structure analysis of the CS and MR genres was undertaken in the Swalesian tradition. The second question was regarding similarities and differences in phraseology between the CS and MR genres. The phraseology of the two genres was explored using Lexical Bundle (LB) analysis. The findings indicate that these two genres differ in pedagogically relevant ways. The third question, which explores the similarities and differences in phrasal complexity between the CS and MR sub-corpora, was investigated using an analysis of word strings. The findings show how genre differences influence the type of linguistic features used to meet the task requirements. The pedagogical implications of this study will be discussed, as EAP practitioners, researchers, students and course book developers need to be aware that situational factors such as genre and discipline need to be considered in characterising student writing at university.

Table of Contents

Certificate of Ethical Approval …….iii

Candidates Declaration Form....iv

Acknowledgement vi

Abstract....viii

Table of Contents....ix

List of Tables xiv

Abbreviations....xvii

Chapter 1 1

Introduction....1

1.1 Background and context....1

1.2 English for Academic Purposes (EAP) 1

1.3 The role of English in education in Oman and other ELF contexts 2

1.4 Situating the current study....5

1.5 Outline of the thesis....8

Chapter 2 9

Literature Review 9

2.1 Introduction....9

2.2 Genre, Genre Analysis and Discourse Community....10

2.3 Classification of student written genres 14

2.3.1 Classification of academic written genres based on surveys....18

2.3.2 Identification of university written tasks based on assignment prompts and course documentation....19

2.3.3 Classification of discipline-specific writing tasks....24

2.3.4 Classification of academic writing tasks based on student writing and other sources....25

2.4 Disciplinary variation in academic writing 30

2.4.1 Studies investigating disciplinary variation in research articles....30

2.4.2 Sub-disciplinary variation in RAs....33

2.4.3 Studies exploring variation across disciplines in written genres other than RAs....34

2.4.5 Corpus techniques to investigate variation in assignments across student genres and disciplines....36

2.4.6 Studies using multi-dimensional analysis to investigate variation in academic written genres 37
2.5 Teaching writing to engineers 40
2.5.1 Engineering vocabulary 43
2.6 Studies of academic writing in the 'Expanding Circle' 44
2.7 Phraseology and Lexical Bundles 46
2.7.1 Identification of Lexical Bundles 50
2.7.2 LB features and their contribution to discourse analysis 51
2.7.3 Biber et al.'s (1999) structural framework for analysing LBs 53
2.7.4 Biber et al.'s functional (2004) framework for analysing LBs 56
2.7.5 Hyland's functional framework for analysing LBs (2008) 59
2.7.6 LBs and disciplinary variation 63
2.7.7 Variation in LB distribution in expert and novice texts 65
2.7.8 LBs in native versus non-native writing 68
2.7.9. Association of LBs with Moves 70
2.7.10 Pedagogical implications 70
2.7.11 Implications of LB studies 72
2.8. Phrasal complexity in academic writing 74
2.8.1 Complex nominals and nominalization 75
2.8.2 Studies of nominal compounds and nominalizations 76
2.8.3 Recent large scale corpus studies on phrasal complexity features 80
2.8.4 Signalling nouns 82
2.8.6 Complex nominals and nominalizations: Complexity or ambiguity? 86
2.9 Chapter Summary 87
Chapter 3 91
Methodology 91
3.1 Civil Engineering writing 91
3.1.1 Context 92
3.2 Data collection methods 93
3.2.1 Corpus compilation 93
3.2.2 Processing of the Corpus 96
3.2.3 Description of the Corpus 98
3.3 Other information sources 100

3.3.1 Course documentation 101
3.3.2 Specialist informants 102
3.4 Categorization of texts into genres 103
3.4.1 Procedure for genre analysis 104
3.4.2 Inter-rater reliability 106
3.5 Corpus analysis tools 108
3.5.1 AntConc Concordancer 108
3.5.2 Sketch Engine 109
3.6 Analysis of Lexical Bundles in the MR and CS sub-corpora 110
3.6.1 Corpus analysis techniques to retrieve LBs in CS and MR assignments 110
3.6.2 Structural and functional classification of LBs 112
3.7 Corpus analysis techniques for extracting and investigating features that contribute to text density 114
3.7.1 Extracting noun strings, nominalisations and attributive adjectives in CS and MR assignments 114
3.7.2 Extracting *the N of (the) N* pattern from the CS and MR sub-corpora 119
3.8 Conclusion 130
Chapter 4 132
Genre Analysis of Case Studies and Methodology Recounts in CE 132
4 Introduction 132
4.1 Identification of Methodology Recounts 134
4.3 Identification of Case Studies 147
4.4 Move Structure of Case Studies in OCAW-CE 149
4.5 Manuals 157
4.6 Site Investigation Reports 158
4.7 Explanations 160
4.8 Exercises 161
4.9 Conclusion 162
Chapter 5 164
Lexical Bundles 164
5 Introduction 164
5.1 Frequency of LBs in the Case Study and Methodology Recount sub-corpora 164
5.1.1 Comparison of LBs in CS and MR sub-corpora 166

5.2 Structure of LBs in OCAW-CE 168
5.2.1 Structural categorization of LBs in CS and MR Assignments in OCAW-CE 168
5.2.2 Less frequently represented structures in the LB lists of CS and MR..... 171
5.2.3 Additional categories..... 173
5.2.4 *Noun phrase with of* fragment, *(PP+) of* phrase fragment, and *other PP* fragment. . 175
5.3 Functions of LBs in OCAW-CE 177
5.3.1 Functional categories of LBs in CS and MR assignments in OCAW-CE 178
5.4 Conclusion..... 190
Chapter 6 193
Nominal Features 193
6.1 Overview of results 194
6.2 Phrasal complexity 195
6.2.1 Nominalisations..... 195
6.2.1.1 Retrospective and Advance Labels 196
6.2.1.2 Nominalisations as technical words 197
6.2.2 The use of nominalisations in OCAW-CE 199
6.2.3 Noun Strings..... 199
6.2.3.1 Types of noun pre-modifiers 202
6.2.4 Discussion on the use of noun strings in OCAW-CE 204
6.2.5 Attributive adjectives 204
6.3 Nouns in *the N of (the) N* pattern 206
6.3.1 Analysis of noun types used by MR and CS texts 206
6.3.1.1 Abstract Nouns..... 209
6.3.1.2 Quantity Nouns 210
6.3.1.3 Place Nouns..... 212
6.3.1.4 Concrete Nouns..... 213
6.3.1.5 Cognitive Nouns..... 214
6.3.1.6 Concrete-Technical Nouns..... 214
6.3.1.7 Gerund..... 215
6.3.1.8 Animate, Group and Proper Nouns 216
6.3.1.9 Range of nouns in the first position of *the N of (the) N* pattern 217
6.3.2 Noun sequences in *the N of (the) N* pattern in CS and MR assignments 218
6.3.3 Discussion of the noun categories in *the N of (the) N* 222

6.4 Conclusion 224
Chapter 7 225
Conclusion 225
7.1 Aim of the study 225
7.2 Summary of the study 226
7.3 Limitations 232
7.4 Future directions 233
7.5 Pedagogical implications 235
7.5.1 Genre pedagogy 235
7.5.2 Teaching applications for Lexical Bundles 237
7.5.3 Teaching linguistic features that contribute to density 239
7.5.3.1. Data-driven learning 240
7.6 Significance of the study 242

List of Tables

Table 2.1 Swales' CARS Model (Swales 2004: 230, 232) 12
Table 2.2 Academic written genres based on survey research and course documentation 18
Table 2.3 Mapping of the BAWE and MICUSP taxonomies 29
Table 2.4 4-word bundles in the Hyland Corpus (2008) using the framework of Biber et al. (1999: 996) 55
Table 2.5 Functional Classification of LBs (Biber, Conrad and Cortes 2004) 58
Table 2.6 Functional Classification of LBs (Hyland 2008) 61
Table 2.7 Summary of studies of phrasal features 77
Table 2.8 SN patterns found in student and published RAs (Aktaş and Cortes 2008) 85

Table 3.1 Number and distribution of texts across the semesters of study 99
Table 3.2 Frequency and dispersion threshold for extracting Lexical Bundles 112
Table 3.3 Concordance lines with management as the node word from the BAWE Sub-corpus of Engineering Case Study assignments 116
Table 3.4 CQL used to retrieve noun strings of different lengths from CS and MR assignments 119
Table 3.5 Examples of concordances where the head noun is outside the node term 121
Table 3.6 Noun category framework for classifying nouns in CS and MR assignments 123
Table 3.7 Semantic noun categories framework (Biber 2006) 123
Table 4.1 Move structure of the Introduction sections of MR assignments in OCAW-CE 138
Table 4.2 Move structure of Methods sections in MR assignments in OCAW-CE 141
Table 4.3 Move structure of Results sections in MR assignments in OCAW-CE 143
Table 4.4 Move structure of Discussion sections in MR assignments in OCAW-CE 145
Table 4.5 Move structure of Conclusion sections in MR assignments in OCAW-CE 146
Table 4.6 Move structure of Introduction sections in CS assignments in OCAW-CE 152
Table 4.7 Move structure of Case Study description sections in CS assignments in OCAW-CE 154
Table 4.8 Move structure of Findings sections in CS assignments in OCAW-CE 155
Table 4.9 Move structure of Conclusion sections in CS assignments in OCAW-CE 156
Table 5.1 Proportion of LBs in the CS and MR sub-corpora 165
Table 5.2 Comparison of most frequently occurring LBs in the CS and MR sub-corpora (raw frequencies) 166
Table 5.3 Structural categories of the LBs in the CS and MR sub-corpora of OCAW-CE with proportions in each corpus 170
Table 5.4 LBs with (NP+) of structures in CS texts fulfilling the discourse functions of location, quantity and quality 177

Table 5.5 Proportion of LBs by main functional categories in CS and MR texts.....179
Table 5.6 Distribution of LBs by sub-function in the CS and MR sub-corpora 180
Table 5.7 Location bundles in the CS sub-corpus of OCAW-CE.....184
Table 6.1 Overview of results: The linguistic devices used to achieve density in the CS and MR assignments 194
Table 6.2 Examples of noun strings of various lengths retrieved from the CS and MR sub-corpora.....201
Table 6.3 Frequencies of N1 categories in the CS and MR sub-corpora 208
Table 6.4 Normalised frequencies of mass, weight, and volume in the CS and MR sub-corpora.....210
Table 6.5 Noun strings with the pattern *the Quantity Noun of the Concrete Noun* in MR assignments 211
Table 7.1 LBs with the pattern *the N of the* followed by concrete nouns……239
Table 7.2 Concordance lines containing the phrase length of within *the N of (the) N* pattern…….243

List of Figures

Figure 5.1 Distribution of LBs by sub-function 1812
Figure 6.1 Noun pre-modifiers in MR and CS assignments in OCAW-CE……...2001
Figure 6.2 Semantic categories of N1 in the MR and CS sub-corpora (normalised frequencies) 2078
Figure 6.3 Percentage of N1 categories in the pattern *the N of (the) N*.....2178
Figure 6.4 Distribution of shared noun-sequences in MR and CS assignments in the N of (the) N pattern 2190
Figure 6.5 Distribution of noun sequences in the N of (the) N pattern in MR and CS assignments 2221

Abbreviations

BAWE	British Academic Written English
CARS	Create A Research Space
CE	Civil Engineering
CS	Case Study
EAP	English for Academic Purposes
EGAP	English for General Academic Purposes
ESAP	English for Special Academic Purposes
EFL	English as a Foreign Language
ELF	English as a Lingua Franca
IMRD	Introduction, Methodology, Results and Discussion
LB	Lexical Bundles
MICUSP	Michigan Corpus of Upper-Level Student Papers
MR	Methodology Recount
OCAW-CE	Omani Corpus of Academic Writing-Civil Engineering
RA	Research Article
SFL	Systemic Functional Linguistics

Chapter 1

Introduction

1.1. Background and Context

Globalization has led to the surge in the number of international students in Western universities as well as in the thousands of satellite campuses worldwide that are affiliated to their Western counterparts. English is the language of instruction and assessment in these remote English as a Foreign Language (EFL) campuses. Many students who now enter university in these contexts, however, are not equipped with adequate academic writing skills. These trends in higher education are reflected in countries such as Oman, where English is used as the lingua franca (ELF), in the manner described by Firth (1996) and House (2012).

Firth (1996: 20) defined ELF as 'a contact language between persons who share neither a common native tongue nor a common national culture and for whom English is the chosen foreign language of communication'. House (2012: 187) echoed Firth's emphasis on the use of English between speakers belonging to different cultures and mother-tongue backgrounds. She points out that 'ELF provides a kind of 'global currency' for people from a great variety of backgrounds who come into contact with one another and use the English language as the default means of communication'.

1.2. English for Academic Purposes (EAP)

In its early stages, in the 1970s and 1980s, writing research in English for Special Purposes (ESP) was influenced by the communicative language teaching revolution, with Widdowson as a leading figure. Following on from this it became popular to take a genre-based approach to EAP research and pedagogy, as exemplified in the work of Swales (1990). The development of Corpus Linguistics led to a further 'paradigm shift in EAP research' (Flowerdew 2016: 16), and social, cultural and departmental contexts also began to be investigated to inform the interpretation of corpus data (Hyland 2004, 2012, Gardner 2008, Nesi and Gardner 2012).

English for Academic Purposes (EAP) research and pedagogy has been described as a social approach; relating to learners' 'process of becoming socialized into specific discourse communities, or communities of practice, communities whose purposes in writing are constrained by institutional, societal, and cultural factors' (Flowerdew 2016:6). In this sense, EAP is intended to facilitate the academic success of students in higher education institutions in specific contexts. EAP can be further categorized into English for General Academic Purposes (EGAP) and English for Specific Academic Purposes (ESAP) (Blue 1988). EGAP is primarily concerned with the teaching of a common core of language devices to students belonging to all fields of study, while ESAP targets students belonging to particular disciplines.

Much of the EAP research on writing tends to be ESAP-oriented language description of texts in specific disciplines, or their variation across disciplines. In this sense, the study described in this thesis has also adopted an ESAP approach as it focuses on the description of texts written by Engineering undergraduates in Oman.

1.3. The role of English in education in Oman and other ELF contexts

The Sultanate of Oman, the country where the research site of this study is located, has participated in global developments in education mainly due to the progressive philosophy of the leader of the country, His Majesty Sultan Qaboos Bin Said. The higher education sector in the country has witnessed an exponential growth over the years; the first public university was established in 1986 and now there are 56 private and public colleges and universities (Oman Academic Accreditation Authority 2019) in the country. Apart from the five universities, most of these higher education institutions are satellite campuses that function in academic partnerships with Western university affiliates; the English language has thus been adopted as the official lingua franca of Oman, and English is used both as the medium of instruction in higher education, and as a means of communication in the world of business where there is a large expatriate workforce (Al Mahrooqi and Denman 2017). English language proficiency is therefore crucial not only for successful academic performance, it is also essential for success in the professions (Al-Jadidi & Sanguinetti 2010). The ELF used in these contexts is not inadequate in any sense, because it possesses a range

of functional and linguistic devices to realize the required communicative purposes. ELF is characteristically able to express a 'multiplicity of voices' (House 2012: 188), and this means that local Omani varieties of ELF may develop in accordance with the national, cultural, and L2 factors influencing its users.

The role of English in Oman can perhaps be understood in terms of Kachru's model (1985) of 'World Englishes'. Kachru's model, consisting of three concentric circles, visualizes the role and functions of English in the various countries in which it is used. The 'inner circle' consists of contexts where English is spoken as the first language, the 'outer circle' consists of countries where English is used as a second language because of its historical roots, and the 'expanding circle' includes contexts where English is a foreign language and is used for the purposes of business or higher education. Although, at the time Kachru (1985) developed his model, English in the 'expanding circle' might have been confined to a 'foreign language' role with limited functions in the countries placed in the 'expanding circle' (e.g. China, Japan, Russia), we find that 'fresh interpretations of changing identities of users and uses of Englishes across the Three Circles' (Kachru, Kachru and Nelson 2006: xviii) are being made.

The 'Expanding Circle' is a large and disparate group in terms of familiarity with and proficiency in the English language, and even in other Arabic-speaking countries such as Syria or Algeria English language use may be somewhat different from that of Oman or the United Arab Emirates.. The differences may be due to various factors such as national policies and the opportunities people have to converse with other speakers of the language. Taking into account the differences among countries in the 'Expanding Circle', we can attempt to bring fresh interpretations to the model by reserving a specific space in the 'expanding circle' for countries located in the Gulf region such as Oman and Dubai, which have a large expatriate work force and where the language of assessment and instruction in higher education is English. In these countries, the language has attained the role of a lingua franca (Al Mahrooqi and Denman 2017).

Particular features of these countries are that the participants in English language communication speak other first languages, and their first languages (L1s) are the medium of instruction in schools but English is adopted as the medium in higher education. Speakers of English in these areas may not need to use as much English or write as elegantly as in the L1 English countries or the users in the 'outer circle', nevertheless, they need to be able to write precisely and clearly in areas where English has a prominent role, such as in business and higher education. This space might be shared with other countries in the 'expanding circle' in which English has a similar role.

Students in these countries entering higher education therefore need to be supported during the course of their undergraduate programmes to attain the grammatically complex and highly specialized discourse of their disciplines. These students need to be able to fulfil the genre requirements of their disciplines by writing clearly and accurately, especially in disciplines such as Civil Engineering where lack of clarity can lead to liability issues and grave financial implications (Conrad 2016).

It can be assumed that students in these contexts have unique needs for academic writing support as they are not similar to students in L1 or 'outer circle' contexts. Unfortunately, to the best of my knowledge, little research on academic writing has been conducted in the countries in the 'expanding circle' which share these contextual characteristics. This type of research needs to focus on the disciplinary writing requirements of students, as a general English approach might not be suitable where students need to attempt disciplinary assignments in unfamiliar genres during the relatively short span of their higher education. Adopting a discipline-specific research approach could narrow down the lexis and syntax to be learned by these students, thereby increasing their chances of writing successful assignments. ELF-using 'expanding circle' countries have hundreds of universities and colleges where students are required to be familiar with the academic writing conventions of English. Research in disciplinary writing conventions is therefore required to meet the unique national, institutional and departmental requirements of these learners, who mostly require a somewhat different kind of academic writing support to those in the 'inner' and 'outer' circles. There is consensus in the research community that the discourse of each

discipline is distinct, and that this has to be acquired by students aspiring to be participants in a 'discourse community' (Gardner, Nesi and Biber 2018, Hyland 2012, Swales 1990). As Berkenkotter et al. (1991:19) affirm, 'students entering academic disciplines need a specialized literacy that consists of the ability to use discipline-specific rhetorical and linguistic conventions to serve their purposes as writers'.

Despite the fact that Omani children are taught English for twelve years at school, the language skills of school graduates entering university continues to be inadequate (Al Mahrouqi et al. 2015). Problems with students' lack of readiness to enter university have led to the establishment of a mandatory one-year preparatory programme in all the higher education institutions in Oman. The learning outcomes of this programme (known as the General Foundation Programme) are prescribed by the Ministry of Higher Education, Oman, and it is intended to prepare students to meet the requirements of their disciplines in terms of General English and basic Mathematical, Computing and Study Skills. However, even though an IELTS Band 5 or an equivalent level of proficiency in English is required for students to enrol on undergraduate programmes in Oman, most students struggle to attain the grammatically complex and highly specialized discourse of their disciplines.

It is in this context that various support mechanisms have been devised by Omani HEIs. However, although many language departments in colleges and universities in Oman offer language modules alongside the content (discipline) modules (Al Jardhani 2017), they often tend to be general in their scope and therefore inadequate as a means to support students' disciplinary writing needs. This is hardly surprising since a discipline-specific approach to teaching and learning language needs to be research-informed and there has been little attention paid to this critical area, as stated earlier.

1.4. Situating the current study

The research site of this study is Middle East College (MEC), one of the largest private higher education institutions in the Sultanate of Oman. Situated in the context just described, students at MEC also require academic writing support. Feedback from subject lecturers, students, external examiners as well as student performance have indicated a need to scaffold

the writing of students throughout the course of their study. The Centre for Academic Writing (CAW) was thus established at MEC in 2014 to address this need. As the Director of the Centre, I work closely with the subject lecturers and students to design and implement mechanisms for academic writing support. The first concern of the Centre was to obtain an understanding of the writing tasks assigned to students in the various disciplines before setting out to devise solutions to their writing problems. Although this research project has emerged from the needs of a Writing Centre seeking to enhance its support for the disciplinary writing needs of the students in one institution, its implications may be wider, especially for other countries similar to Oman in the 'expanding circle'.

I was influenced by the research on university student writing undertaken by the British Academic Written English (BAWE) (Alsop and Nesi 2009) and the Michigan Corpus of Upper-Level Student Papers (MICUSP) (Römer and O'Donnell 2011) project teams in the UK and US, respectively. They compiled large systematically archived corpora of proficient university student writing to investigate the types of genres used across the disciplines. Studies using smaller private corpora are also reported in the literature (Hyland 2008), but since none of them, to the best of my knowledge, examine university writing in the countries where English is the lingua franca, I decided to start my research project by creating a corpus of student writing. Creating a specialized corpus can help me obtain an overview of the writing assigned to students in a discipline in terms of the genres assigned, their rhetorical structures, and the linguistic devices used to realise their unique communicative purposes.

The aim of my project is thus to create a representative corpus of Omani university student writing and attempt to characterise the principal assignment types by describing their rhetorical structures and providing a linguistic description of their phraseology. Nesi and Gardner (2012) explain the pedagogical applications of BAWE; the rationale for compiling the Omani corpus of student writing is similar. Descriptions of the language features of Omani student texts will help EAP instructors identify the needs of these students, and will help subject lecturers think about the purpose of their assignments and their own expectations of these students. My research can also add to the literature on this strand of ESAP which other researchers can build upon. Omani student assignments might differ from assignments

written by native speakers because of institutional, departmental, cultural and other factors. Describing high quality assignments written by native speakers from another context might therefore not serve the purpose of this study. The assignments included in the Omani corpus of student writing have been chosen by subject lecturers themselves as meeting their expectations. Hence I conclude that they are 'fit for purpose', even if this does not necessarily mean that they would be considered proficient in other contexts removed geographically and culturally from Oman.

In terms of the scope of my thesis, I will focus on one discipline because this corpus design will enable me to examine large amounts of student writing within a single discipline and isolate the effects of genre. I chose to investigate student writing in Civil Engineering (CE) for two reasons: the importance of writing for Civil Engineering students and practitioners (Conrad 2018) and the dearth of discipline-specific learning resources. Although Conrad et al. (2016) provide some resources to improve the writing of CE students, they are limited to some workplace genres and some general and stylistic linguistic items such as tips on avoiding ambiguous writing and the use of active sentences to emulate the writing of CE practitioners. Other resources available for students cater for engineering disciplines in general, for example, Engineering Academic Formulas List (Fox and Tigchelaar 2015), and Massive Open Online Courses (MOOCs) on technical writing for engineers hosted by universities. Although these resources as well as others on specific linguistic features (Vincent and Nesi 2018) can help students in improving their general and specific English skills, they need to be complemented with knowledge of the specific linguistic devices and macrostructure of genres assigned to students as coursework. In spite of the resources currently available, students still face very great challenges when attempting unfamiliar genres in a foreign language.

There have been few attempts, to the best of my knowledge, to investigate the types of pedagogical genres assigned to CE students over the course of their study. Identifying these 'occluded' genres and describing them linguistically is the first step towards understanding the writing requirements of students and designing appropriate learning materials in order to help them meet their disciplinary writing needs. The term 'occluded'

originally applied by Swales (1996: 46) in this context, has been defined by Kanoksilapatham (2012: 297) as 'a genre to which access is normally denied to those outside the discourse community.' Therefore, my study will be the first to characterise student writing in Civil Engineering in the 'expanding circle' based on a representative corpus of student writing and an understanding of the contextual factors that shaped these texts.

1.5. Outline of the thesis

The thesis has seven main chapters. Chapter Two reviews studies on student genres, the literature on disciplinary variation in student writing and relevant studies in phraseology and phrasal complexity that contextualize my analyses. The third chapter explains the methodology used for identifying the genres in my corpus, the corpus linguistic techniques used to retrieve the Lexical Bundles identified in this study and the patterns and linguistic devices that contribute to the lexical density of the assignments in my corpus (henceforth the Omani Corpus of Academic Writing-Civil Engineering (OCAW-CE)). The methods used to analyse the data thus retrieved are also explicated along with rationales for their choice. The fourth chapter contains the findings from the genre categorization procedure and discusses these findings. It provides a detailed description of the two most important genres identified in the corpus, along with a brief description of the other genres that were identified, using the Swalesian text analysis method. Chapter Five presents the results of the analysis of the Lexical Bundles retrieved from the two sub-corpora and a comparison of their use in the two genres. Chapter Six examines the linguistic devices and patterns that contribute to phrasal complexity along with an investigation of the semantic noun categories favoured by the two genres within the pattern *the N of (the) N*. Chapter Seven concludes by exploring the pedagogical implications of my findings and discussing applications for practice.

Chapter 2

Literature Review

2.1. Introduction

The purpose of this chapter is to introduce the literature related to genre analysis, phraseology and phrasal complexity features in academic writing. These three perspectives are combined to describe the two main genres in the Civil Engineering corpus designed for this study. In this chapter, I first discuss Swales' influential genre analysis approach (1990, 2004) which is based on the demarcation of moves and steps in a text and their association with the communicative purpose of the genre. I then examine the notion of phraseology by investigating Lexical Bundles (Biber et al. 1999, Hyland 2008). The chapter also contains a review of research on phrasal complexity features which are increasingly being acknowledged as important characteristics of academic writing (Biber and Gray 2016, Gardner, Nesi and Biber 2018). The phrasal complexity features I particularly focus on are noun strings, nominalizations and attributive adjectives.

Section 2.2 discusses the concept of genre, genre analysis and the associated notion of 'discourse community' as an understanding of these is vital to the review of the literature on student writing and disciplinary variation. Section 2.3 provides a broad overview of attempts to develop taxonomies of student writing genres. This discussion will justify the approach adopted in this study to identify and classify the student academic CE genres at MEC. Section 2.4 deals with studies that have investigated disciplinary and genre variation through various methods including genre and corpus analysis techniques. The aim of this section is to trace the increasing emphasis and evidence in the literature on the specificity of academic discourse. The review of studies on disciplinary and genre variation here is relevant for this study, which seeks to uncover the similarities and differences between the two main genres in a single discipline. A detailed examination of the BAWE project is included in this chapter since, like the current study, it examines authentic student writing and is in fact the inspiration for this one. Studies in engineering writing, and academic writing in some areas of the 'expanding circle' feature in Sections 2.5 and 2.6. Section 2.6 serves to establish the

under-researched status of academic writing in countries in the ‘expanding circle’ where English is used as the lingua franca. Section 2.7 discusses the frameworks developed to investigate the structure and functions of Lexical Bundles along with studies that are based on these. The eighth section of this chapter focuses on noun strings, nominalizations and attributive adjectives, as phrasal complexity features.

2.2. Genre, Genre Analysis and Discourse Community

‘Genre’ is a fundamental concept underlying research in academic writing. Although it is a difficult term to define precisely, different schools of genre analysis have reached a consensus on some of the characteristics of a genre. Martin (1997: 13) defined it as ‘the system of staged goal-oriented social processes through which social subjects in a given culture live their lives’. Swales (1990: 46) defines it as ‘a class of communicative events, where communicative events comprise not only the discourse itself and its participants, but also the role of that discourse and the environment of its production and reception, including its historical and cultural associations’. Miller (2000: 8) regards genres as ‘solutions’ to ‘recurring rhetorical problems’ which gain ‘institutional force’ through repeated use. Likewise Devitt (2004: 31) states that a genre is ‘a reciprocal dynamic within which individuals' actions construct and are constructed by recurring context of situations, context of culture, and context of genre’.

These descriptions suggest that genres have a social purpose, that they have distinct stages and that there are communities sharing preferred genres which are used as a means of communication amongst their members. It follows that the social purpose, audience, context and the role of the writer help to identify genres and that there are distinct lexico-grammatical features that realise the functions of the rhetorical stages of a text (Nesi and Gardner 2012: 25).

Genre analysis in the ‘Swalesian’ tradition is one of the most frequently used approaches for investigating textual features and rhetorical patterns of written texts to uncover disciplinary variation. According to Swales (1990, 2004), all written genres contain certain obligatory and optional ‘moves’ and each ‘move’ may also have one or more ‘steps’ which can be

arranged in a certain sequence (see Table 2.1). Since the publication of Swales' 'Creating a Research Space' (CARS) Model (1990), the Research Article (RA) has been the most studied academic genre and CARS has been the most influential textual analytical tool used in the analysis of RAs (Anthony 1999, Basturkmen 2012, Lin and Evans 2012). Swales applied genre analysis to RA introductions to demonstrate how 'moves' and 'steps' realise their intended communicative functions. The three obligatory moves in RA Introductions identified by Swales (2004) are (1) establishing a territory; (2) establishing a niche; and (3) presenting the present work.

The CARS model is flexible, as the notion of 'cyclicity' of moves or recurrence of moves is acknowledged. For example, the model allows for the 'reviewing literature' step to recur 'throughout the introduction and indeed throughout the article as a whole' (2004: 227). The possibility of variation among disciplines is suggested by the inclusion of a 'PISF (probable in some fields)' option (see Table 2.1). For example, to accommodate this variation, Move 3, which is labelled 'presenting the present work' and includes one obligatory step, also includes three optional steps and three 'probable in some fields' steps (Swales 2004: 230).

Table 2. 1. Swales' CARS Model (Swales 2004: 230, 232)

Swales' general approach, involving moves and steps, has been adopted for the analysis of other sections of RAs (Basturkmen 2009, 2012, Kanoksilapatham 2005, 2015, Samraj 2005), and for pedagogic genres such as Methodology Recounts and Case Studies (Nesi and Gardner 2012, Parkinson 2017, Tribble and Wingate 2013).

Closely associated with the concept of genre is the idea of 'discourse community'. Herzberg (1986: 1) claims that

> discourse operates within conventions defined by communities… language use in a group is a form of social behaviour….. discourse is a means of maintaining and extending the group's knowledge and of initiating new members into the group, and… discourse is epistemic or constitutive of the group's knowledge (Herzberg 1986: 1).

Herzberg's statement implies that there is a unique discourse shared among groups of individuals belonging to a community which holds knowledge in common, and that new members need to learn this discourse to join this 'discourse community'.

Swales (1990) proposed a set of six criteria for membership in a discourse community: shared goals, communication among members using a set of genres, interaction among the members, preferred genres belonging to specific discourse communities, use of

specialized and technical lexis, and an ideal ratio of experts to novices. This accords with the claim by North (2005) that 'as part of their socialization into a disciplinary area, students also assimilate the literacy practices accepted within that discipline'. These researchers suggest that it is possible to identify distinct communities that share discourse conventions and social goals, and which have at their fringes a group of novices who can only aspire to enter these communities if they are familiar with its preferred genres and lexis, and if they have interactions with community members.

These descriptions of 'discourse communities' may not completely apply to the university context of teachers and students. To do so would entail that students, after initiation into their discourse communities by their teachers who are the 'experts', would eventually themselves gain enough expertise in the preferred genres to become fully-fledged members. However, this is probably an over-simplification as, more often than not, students never reach the level of expertise required of academic discourse communities. This could be due to many reasons: they might not regard academic interaction as their eventual goal, they might not wish to engage in academic research after they have gained their degree or to continue to interact with members of the academic discourse community, and their eventual workplace might require them to communicate using a new, different set of 'preferred genres'. Nevertheless, while they are studying, students are required to be familiar with the discourse conventions of their disciplines to some extent, in order to produce pedagogic genres or genres that students are required to produce at university, even if these may not necessarily be replicated in workplace contexts. The assignments collected for my corpus were written by students whose work reached the expectations of their lecturers, and who might therefore be considered to be on the fringes of the academic engineering discourse community. It is hoped that analysis of their work will inform the writing support provided for weaker students, so that their interaction with this discourse community might improve, even though they may never become fully-fledged members. The more closely the writing of these weaker students meets departmental requirements, the more likely they will be to succeed in their studies, and eventually find appropriate employment.

Despite some issues surrounding its precise definition and application, the term 'discourse community' is useful to understand the differing epistemologies and social practices of academic disciplines. These are the basis for the broad division of disciplines into 'hard and soft' and 'applied and pure' (Becher 1989, Becher and Trowler 2001, Neuman, Parry, and Becher 2002), and are also acknowledged by other researchers who have probed the possible differences among disciplines (e.g. Braxton 1995; Squires 2005). Based on Neuman, Parry, and Becher's (2002: 406) framework, we can group disciplines into 'hard' and 'pure' (e.g. Chemistry, Physics), 'soft' and 'pure' (e.g. History, Literature), 'hard' and 'applied' (e.g. Engineering), and 'soft' and 'applied' (e.g. Business, Education). They point out that 'hard' and 'applied' knowledge, which is typified by Engineering disciplines, 'derives its underpinnings from hard pure enquiry, is concerned with mastery of the physical environment and geared towards products and techniques'.

2.3 Classification of student written genres

There has been an increasing interest among researchers and practitioners in understanding the writing tasks students are required to complete at university. This attention could partly be due to widening participation at the local level as well as the large influx of international students into the Western university system, many of whom come from EFL contexts and therefore tend to lack the required academic writing skills.

Studies that seek to examine the kind of academic writing that is done by university students can be categorized into three types: those that primarily depend on surveys and task completion prompts, those that examine course documentation such as assignment rubrics and module descriptors, and those that use a combination of course documentation, interviews with subject lecturers, and analysis of authentic student writing. All of these studies attempt to devise taxonomies of assessment tasks in higher education to inform pedagogy, assessment and research. Table 2.2 provides a summary of the studies reviewed in Sections 2.3.1, 2.3.2 and 2.3.3. Sections 2.3.1 to 2.3.4 review studies that identified student assignments at university based on surveys, course documentation, discussions with subject specialists and student writing, respectively.

Study	Source of Data	Details regarding disciplines/size of data set and context	Classification
Horowitz (1986a)	Assignment tasks	• 54 coursework assessment tasks from 17 departments • US (Western Illinois University, a medium sized state university)	1. Summary of/reaction to a reading 2. Annotated bibliography 3. Report on a specified participatory experience 4. Connection of theory and data 5. Case study 6. Synthesis of multiple sources 7. Research Report
Horowitz (1986b)	Examination prompts	• 284 examination prompts from 15 departments • US (Western Illinois University, a medium sized state university) • Same dataset as previous study	1. Display familiarity with a concept 2. Display familiarity with the relationship between/among concepts 3. Display familiarity with a process 4. Display familiarity with argumentation
Hale et al. (1996)	Assignment tasks	• 162 undergraduate and graduate courses • 6 disciplinary areas • 8 US universities	1. Locus of writing 2. Length of product 3. Genre 4. Cognitive demands 5. Rhetorical task

			6. Pattern of exposition
Study	**Source of Data**	**Details regarding disciplines/size of data set and context**	**Classification**
Rosenfeld et al. (2004)	Survey instrument	• 39 task statements in survey administered to 800 respondents • 6 broad disciplinary areas • 30 US universities	Identified important tasks across discipline and level based on ratings of respondents
Moore and Morton (2005)	Assessment tasks	• 155 assessment tasks across 28 disciplines and 20 IELTS Task 2 prompts • 2 Australian universities	1. Genres 2. Information Source 3. Rhetorical Function 4. Object of enquiry
Jackson et al. (2006)	Questionnaire responses with option for faculty comments	• 47 questionnaires • 13 disciplines • 3 African universities	1. Summary of/reaction to readings, 2. Experimental (laboratory) report 3. Research paper 4. Essay
Carter (2007)	Assessment tasks Course documentation Interviews	• 51 programs from 9 undergraduate colleges • US university	1. problem-solving 2. empirical inquiry 3. research from sources 4. performance
Cooper and Bikowski (2007)	Course syllabi Interviews	• 200 course syllabi from 20 academic departments • US Midwest university	1. library research papers 2. reports on experiments/projects 3. article/book reviews 4. plans/proposals, 5. summaries

			6. case studies, 7. unstructured writing 8. journal articles 9. essays 10. annotated bibliographies, 11. miscellaneous (adapted from Horowitz and Hale et al.)
Study	**Source of Data**	**Details regarding disciplines/size of data set and context**	**Classification**
Gillett and Hammond (2009)	Course documentation	• Assessment tasks in 2,367 modules • 5 broad disciplinary areas • UK university	Six categories under which writing tasks could be grouped: Tasks, Medium, Who Assesses, Cognitive, Time span, and work-related and listed 22 features found in them
Melzer (2009)	Assignment prompts	• 2100 assignment prompts from 400 courses • Collected from websites	1. Writer 2. Text 3. Audience 4. Exploratory

Study	Source of Data	Details regarding disciplines/size of data set and context	Classification
Shi and Dong (2015)	Graduate assignments	• 143 assignments • Across 12 faculties • Canadian university	1.scholarly essay 2.summary and response 3.literature review 4.project 5.review 6.case analysis 7.proposal 8.exam 9. creative writing

Table 2. 2. Academic written genres based on Survey Research and Course Documentation

2.3.1 Classification of Academic Written Genres Based on Surveys

Many studies which have overviewed university writing have depended on faculty members' and students' opinions and intuitions using task completion prompts and surveys (Bernhardt 1985, Bridgeman & Carlson 1984, Casanave & Hubbard 1992, Eblen 1983, Trimmer 1985; see Table 2.2). However, the methods used in these types of studies are open to bias and subjectivity as they tend to 'force a set of preconceived categories onto their respondents' (Cooper and Bikowski 2009: 3). For example, Jackson et al. (2006: 277), in their study which aimed to gather information about the types of tasks assigned to undergraduate science students at a South African university, asked lecturers to categorize the writing tasks they assigned into six types: the essay, summaries of one and multiple reading texts, lab work, field trips, and reports on 'data supplied by lecturer'. This limited set of options may not be comprehensive enough to reflect the actual writing tasks and there is the possibility that lecturers mismatched the tasks and the labels provided in the absence of thorough descriptions of these tasks. Nevertheless, these 'survey studies' provide a starting point for the crucial task of attempting to understand the type of writing demanded of students, in terms of the various genres attempted by students and what lecturers value in student writing.

Rosenfeld et al.'s (2004) study is much broader in scope, with participants including faculty members from six broad disciplinary areas (English, Psychology, Education, Life Sciences, Physical Sciences, and Engineering) from 30 universities in the United States. More than 800 faculty members participated in a survey to choose the most important writing tasks assigned at various levels (undergraduate, graduate, doctoral) in the six disciplinary areas. They chose from a list of thirty nine items proposed by Rosenfeld et al. (2004: 49) as being required for 'competent performance in a wide variety of academic areas'. However, the studies of both Jackson et al. (2006) and Rosenfeld et al. (2004) share the drawback of relying on responses to surveys unaccompanied by other data to identify written tasks assigned at university. There is also little flexibility for respondents who are forced to choose among the options available in the survey instrument.

2.3.2. Identification of university written tasks based on assignment prompts and course documentation

An alternative approach for identifying and categorizing tasks assigned to university students consists of inspecting actual examination and assignment prompts as well as course documentation such as module descriptors (Carter 2007, Cooper and Bikowski 2009, Gillett and Hammond 2009, Hale et al. 1996, Horowitz 1986a, 1986b, Melzer 2009, Moore and Moreton 2005, Shi and Dong 2015; see Table 2.2).

The most influential studies of this type were conducted by Horowitz (1986a, 1986b) who first investigated the kind of writing that is assigned to university students and reported the possibility that students might be doing different types of writing in different disciplines. The significance of these studies can be gauged from the number of articles and dissertations which have been built upon Horowitz's work (Cooper and Bikowski 2007, Hale et al. 1996). Some later studies of this type were much larger and more systematic and included multiple institutions and hundreds of prompts. The most significant of these is Melzer (2009) who examined 2100 assignment prompts available online, in 400 courses from 100 post-secondary institutions.

However, the genre categories developed by these researchers tend not to be very transparent, as a detailed description of each assignment is not provided and, moreover, they are not supplemented by additional data sources such as interviews and student written work, with the exception of a few cases (e.g. Carter 2007, Cooper and Bikowski 2009). For example, Melzer (2009) classified the texts according to six functions ('transactional', 'informative', 'exploratory', 'poetic', 'persuasive' and 'expressive') and the audience being addressed ('examiner', 'self', 'peer', 'informed', 'novice', 'wider audience', 'generalized reader'). The functions are rather intuitive and consist of broad and overlapping categories which might not be replicable in other contexts. Melzer (2009) himself draws attention to this feature of his classification system; he speaks about the difficulty of describing the 'Research Report' because of disciplinary variation; in the absence of details regarding the rhetorical sections of the texts belonging to this pedagogic genre, and the lexico-grammatical features employed to realize the communicative functions of each section, the description is incomplete and would make it difficult to identify a similar genre in another context. The methodology used in these studies does not afford such detailed descriptions since the student assignments themselves need to be examined to provide the kind of information required to report details at this level.

Another limitation of these studies is that, in the absence of a consistent and overarching framework based on uniform criteria, there are predictable overlaps in the types of student writing identified in different contexts. For example, Moore and Moreton's (2005) 'information source', Carter's (2007) 'research from sources' and Shi and Dong's (2015) 'literature review' apparently refer to assignments that require students to integrate information from various sources. The confusing nomenclature needs to be addressed by a comprehensive framework that is informed by linguistic descriptions based on student written texts and verified by lecturers and course documentation. These researchers did not perhaps have the means to access other sources of information. For example, the data used in Melzer's study is web-based, and thus he did not acquire access to the assignment developers and actual student writing samples.

These studies have also tried to determine the most frequently assigned types of assessments, although, in the absence of detailed descriptions, they have limited applicability. For example, Melzer (2009) identified the 'term paper' and the 'short-answer exam' as the most frequent types of assessments, Moore and Morton (2005) report that the 'term paper' is one of the most frequently assigned writing tasks, and Cooper and Bikowski (2009) found the most commonly occurring genre to be the library research paper followed by the 'project report' and the 'book review' (the least common to be 'examination essays' and 'short tasks'). Cooper and Bikowski's 'library research paper' seems to resemble the 'term paper' identified by Melzer (2009) and Moore and Morton (2005) but without a thorough description of these genres and the categorization process, we can only attempt to make comparisons. Evaluated on these bases, these studies are therefore not sufficiently rigorous. This highlights the need for a more robust framework based on thorough generic and linguistic descriptions.

Carter (2007) also identified the genres favoured by certain disciplines at nine colleges in a US university but with an emphasis on the nature of these disciplines. He reports, for instance, that the scientific disciplines, which emphasize 'empirical inquiry', favour genres which manifest this feature, such as methodology recounts or project proposals. He draws attention to disciplinary 'ways of doing' by providing examples of genres found in certain disciplines (e.g. 'problem-solving' in Textile Engineering). What is unique about Carter's study is that faculty members themselves were asked to reflect on what learning outcomes they expected students to achieve and then identify the genres most suitable for their discipline to assess these outcomes. Carter's (2007: 385) is an insightful study that attempts to explain the nature of different disciplines and show how pedagogic genres are chosen by subject lecturers to assess students according to disciplinary ways of 'knowing, doing and writing'. The approach taken by Carter (2007) suggests that a way of taking this type of research forward, is to develop a better understanding of genres in specific disciplines by inviting subject lecturers to identify and reflect on their assessment choices. A feasible approach towards conducting this type of study would be to focus on one institution, especially if resources are scarce.

Other studies discussed in this section (Hale et al. 1996, Melzer 2009) aimed to identify pedagogic genres across a range of universities and courses by examining course documentation and assignment prompts. As we might expect, it is not possible for such datasets to be fully complete due to the breadth of their scope. These studies have, however, partially fulfilled their objectives, which were to gain an understanding of the writing requirements across an entire country or region so as to generalize to wider contexts. The disadvantage of this approach is that although the scope is wide, involving a large number of tasks, courses and institutions, these studies lacked the resources to provide a thorough linguistic analysis of assignment types.

An alternative to this broad approach to examining student tasks would be to focus on the tasks assigned in a discipline in a single institution, so that there is a stronger claim to be representative, albeit of a smaller context. Since there would be less data, they could be looked at in greater depth, and linguistic descriptions of the genres including their structure, phraseology and other lexico-grammatical features could be examined. This approach would involve a comprehensive sweep of all the genres assigned as coursework in a specific institution in one discipline, and would be capable of informing pedagogy and writing support provision in the institution because the pedagogically relevant genres would all have been identified. The findings from such studies might also be generalizable to a wider context. For example, studies of this type conducted in an institution in Oman might be relevant to similar institutions in countries occupying the same area in the 'expanding circle', where English is the lingua franca. Although analysing student texts from a single institution may be seen as a limitation, individual researchers may not have the resources to investigate multiple contexts, and are certainly best placed to discover the disciplinary requirements of their own specific working environment. Synthesizing the findings of studies situated in single institutions can enable us to discover similarities and differences between student texts produced in different contexts.

Some studies (Gillett and Hammond 2009, Melzer 2009) that have attempted to identify university writing tasks by examining course documentation and prompts have developed frameworks that not only identify genres but also include other contextual

variables. For example, Gillett and Hammond (2009) developed six dimensions described as 'process-oriented', under which tasks are grouped: 'Tasks' (e.g. Role play, open book), 'Medium' (oral, numeric, diagram), 'Who Assesses' (e.g. self, peer), 'Cognitive' (intellectual processes), 'Time span' (e.g. portfolio, periodic), and 'Work-related' (simulation of work place genres). The emphasis of this type of study is more on the processes involved in assessment, to support assessment developers and writing support tutors, rather than on the identification of different types of writing. These studies have provided additional perspectives on assignment writing and might therefore fall into the category of critical English for Academic Purposes (EAP) studies which attempt to influence pedagogy.

To sum up, the studies in this section reveal a range of pedagogic genres used in assessment and report variation in the predominance of certain kinds of assessments in different disciplines. One limitation of these studies is that the categories are not extensively described and therefore have low reliability; researchers adopting these classification systems are likely to be left with some unanswered questions when attempting to group texts from another context under these categories. The various genre labels attached to texts without extensive descriptions add to the confusion; this impedes comparison across studies. Moreover, without access to student writing, predicting the response to these tasks would involve 'an act of interpretation on the part of the analyst' (Moore and Morton 2005: 46). The main limitation in all these studies is that teachers were not interviewed and student writing samples were not examined, drawing the concession from Hale et al. (1996: 3) that there was 'a good deal of inference in classifying the assignments'.

Hence it can be concluded that none of these frameworks is comprehensive and flexible enough to classify all the types of student writing that may be encountered in other contexts because insufficiently rigorous attempts were made to group assignments based on common purpose, audience and rhetorical sections. These studies also have limited significance because they do not appear to have directly informed pedagogy or assessment, in the form of materials development or the design of coursework.

2.3.3 Classification of discipline-specific writing tasks

The discipline specificity of written genres is being increasingly acknowledged in EAP academic writing research, as can be seen by the studies reviewed in this section, which focus on a single discipline. The advantage of studies that have used this approach is that conclusions are based on a substantial amount of data from a single discipline and a single institution.

Canseco and Byrd (1989) examined the syllabi of 48 business courses at a south-eastern university in the US. They arrived at seven types of assignments: 'exams', 'written responses to problems', 'projects', 'case studies', 'papers', 'reports', and 'miscellaneous assignments'. However, the nomenclature for these assignments were adopted from the documents examined and may not reflect the actual expectations of the task, especially as detailed descriptions for each coursework were not provided. It is not clear, for instance, how a 'project' differs from a 'report'.

Zhu's (2004) study examined the types and characteristics of assessment tasks given to business students by studying 95 course syllabi and some handouts and 12 student writing samples, as well as by interviewing six faculty members. Zhu identified nine pedagogic genres out of which three were found to be the most frequently assigned: 'case analysis', 'article/book report' and 'business report'. The rhetorical sections of each genre were briefly described, for example the inclusion of an executive summary and 'justification of the design choices' in a 'Design Project' (Zhu 2004: 122). This discussion of the rhetorical sections shared across texts belonging to a single genre or what can be called 'staging' of texts was later taken up by Nesi and Gardner in much more detail (2012). Zhu's (2004) study also has a few limitations: very few (12) assignments are represented in the analysis, and tasks like 'take-home assignments', 'exercises' and other texts which did not demand much writing were excluded. Nevertheless, the methodology adopted by this study is an improvement over the others reviewed until now because of the multiple data sources used (course syllabi, interviews, student writing) and the attempts made to describe the structure of the texts belonging to a genre. By restricting the data collection to a single discipline and institution,

this type of study can provide a more in-depth characterization of student assignments because of the manageable amount of data. Institutions in similar contexts can benefit from the findings of such studies. Clearly, these studies demonstrate a different approach to examining and identifying student genres from those discussed in the previous two sections.

Sections 2.3.1, 2.3.2 and 2.3.3 have discussed attempts to identify and categorise student genres at university. A comprehensive understanding of student writing requirements, however, can only be obtained by examining corpora containing a substantial number of authentic student assignments.

2.3.4 Classification of academic writing tasks based on student writing and other sources

The main focus of this section is on studies that have based their classification of university writing tasks on systematically designed corpora consisting of student texts. A corpus is defined as a collection of texts in electronic form 'selected according to external criteria to represent, as far as possible, a language or language variety as a source of linguistic research' (Sinclair 2005: 16). As pointed out by Biber et al. (1998: 177), 'representativeness in corpus design is a crucial concern for all corpus-based studies, particularly in seeking a comprehensive description of language variation'.

There are comparatively few corpora of student coursework perhaps because of the difficulty of collecting large samples of such 'occluded' genres as academic assignments. Corpora of student or pedagogic genres in the disciplines need to be distinguished from learner corpora which mainly contain essays produced by EFL and ESL learners in the language classroom or under examination conditions. Unlike learner corpora, for which it is comparatively easy to collect texts, there are problems associated with collecting large numbers of assignments written in different university disciplines, especially if the compilers are dependent on the students themselves donating their work.

Studies that have developed classifications of student written genres based on analysis of corpora of student work represent a major advance compared to those reviewed

in the previous three sections (2.3.1, 2.3.2 and 2.3.3). They have demonstrated how it is possible to conduct genre analysis and quantitative and corpus-based analyses of linguistic features and compare and contrast the findings across genres and disciplines. I will now discuss the British Academic Written English (BAWE) corpus and the Michigan Corpus of Upper-level Student Papers (MICUSP), the first systematic and large-scale corpora of student academic writing.

The aim of the BAWE project was to provide an overview of university student genres in the UK. This project involved the collection of proficient student writing at undergraduate and Master's level from four UK HE institutions. Student writing samples were collected from 30 disciplines in four broad disciplinary areas across four levels of study, and an effort was made to obtain a balanced distribution and representativeness of texts across these disciplines and levels. The resulting corpus contained over 6.5 million words (Alsop and Nesi 2009) and it has been widely used by researchers, EAP tutors and students (Durrant 2014, Vincent and Nesi 2018).

A robust genre classification system was devised by Nesi and Gardner (2012) based on the shared purposes and the rhetorical staging of texts and informed by multiple data sources such as student assignments, course documentation and discussions with subject lecturers. Assignments were grouped into 13 genre families on this basis: Explanations, Exercises, Critiques, Essays, Research Reports, Literature Surveys, Methodology Recounts, Problem Questions, Proposals, Design Specifications, Case Studies, Narrative Recounts, and Empathy Writing.

These 13 genre families were then subsumed under five broad functions of academic writing: (1) demonstrating knowledge and understanding; (2) developing powers of independent reasoning; (3) building research skills; (4) preparing for professional practice; and (5) writing for oneself and others. These functions are termed 'social purposes' by Nesi and Gardner (2012: 26). Texts were grouped under each of the broad functions 'at a level of delicacy that groups similar genres across disciplines' on the basis of their shared social purposes and staging (Nesi and Gardner 2012: 29). For example, the 'product evaluation'

and ‘book review’ genres are grouped under the Critique family, even though they are typically associated with different disciplines, because both have a similar social function. Another example is the Company Case Study type of assignment in Business and the patient Case Study report in Medicine, grouped under the ‘Case Study’ genre in spite of the differences between their objects of study. This type of categorisation cuts across disciplinary boundaries to trace the common functions running through genres across disciplines. It provides researchers with a classification scheme which can be applied to all disciplines and genres of student writing irrespective of context.

Another important enhancement over previous studies is the systematic labelling of pedagogic genres in the BAWE corpus, achieved despite the obscure assignment labelling practices followed by various academic departments. The nomenclature system used in the BAWE corpus is more transparent than those reviewed earlier in this section because detailed explanations of the characteristics of each genre are based on textual analysis, documentary evidence, and interviews with content teachers. The use of such complementary data sources in BAWE ensured that ‘text and context’ were both considered during the categorization process (Nesi and Gardner 2012). The identification of each genre is also based on purpose, staging, and corresponding lexico-grammatical features and provides researchers with a useful model that can be applied to other contexts.

Although it is smaller in scope than BAWE, another noteworthy corpus project is the Michigan Corpus of Upper-level Student Papers (MICUSP), which is based on student writing collected at the University of Michigan in the US (Römer and O’Donnell 2011). MICUSP, like BAWE, is a collection of proficient student writing in different disciplines and at different levels of study (final year undergraduate, first, second and third year graduate). It comprises around 830 papers that were collected from 16 disciplines through an online submission system. The majority of the assignments in MICUSP belong to the Arts and Humanities and the Social Sciences.

The MICUSP team did not adopt existing classifications of student writing and instead arrived at a taxonomy of student genres based on a data-driven approach, much like

the BAWE project. The texts were analysed by linguists, EAP teachers and students and seven 'text types' were finally arrived at: 'Argumentative Essay', 'Creative Writing', 'Critique/evaluation', 'Proposal', 'Report', 'Research Paper' and 'Response Paper'. However, since, unlike Nesi and Gardner (2012), there is a lack of a broad overarching categorization framework based on the purpose, audience and staging of texts (macro-structure), this categorization is somewhat ambiguous.

The BAWE and MICUSP projects share some similarities. Both include metadata about the background of the contributors and the text and are among the first attempts at compilation of systematic corpora of student academic writing. BAWE allows users to apply corpus query software and searches can be based on discipline, level of study and genre. BAWE is much larger in scope, comprising six and a half million words (MICUSP is about 2.6 million words), and it is publicly archived and available to other researchers. Table 2.3 is an attempt to match the BAWE and MICUSP taxonomies. In some cases MICUSP genres do not map precisely onto the BAWE genres. At the same time, there is some association among some of the genres identified by the BAWE and MICUSP projects; five genres, the Critique, Essay, Research paper, Proposal and Narrative Recounts seem comparable. This mapping demonstrates the more inclusive nature of the taxonomy devised by Nesi and Gardner (2012). It may also indicate that student writing in the UK is more diverse than in the US, or that the collection methods of MICUSP may have been less effective. Nesi and Gardner's (2012) framework, nevertheless, appears to be more readily applicable to other contexts than the others discussed in this review.

BAWE	**MICUSP**
1. Explanation	
2. Exercise	
3. Critiques	Critique/evaluation
4. Essays	Argumentative
5. Research Reports	Report/Research paper
6. Literature Surveys	

7. Methodology Recounts	
8. Problem Question	
9. Proposal	Proposal
10. Design Specification	
11. Case Study	
12. Narrative Recounts	Creative writing Response paper
13. Empathy Writing	

Table 2. 3. Mapping of the BAWE and MICUSP Taxonomies

The model of the BAWE project in terms of the methods used to collect data and the analysis of student assignments seems to represent a relatively systematic and effective approach to investigating student genres. The BAWE corpus of student writing is complemented by contextual information sources such as the views of specialist informants and course documentation. Although this combination of using corpora of actual student writing along with other contextual information seems to be a more comprehensive approach towards identifying and classifying student writing genres in one context (UK), the BAWE approach might need to be adapted for other contexts and also for studies that seek to provide detailed linguistic characteristics of genres within a discipline. BAWE represents student writing in the UK and might not necessarily describe the pedagogic genres produced in other contexts which are shaped by contextual variables such as institutional and departmental requirements. The necessity of creating corpora representing specific contexts is therefore crucial for an understanding of the pedagogic genres and writing requirements in these unique contexts. I have also pointed out the need to create corpora of single-discipline texts for thorough linguistic descriptions of genres based on an adequate number of texts. This would be another way of making a study comprehensive, representative, and of pedagogic value, although smaller in scope than large-scale studies such as the BAWE project.

The studies reviewed above help to uncover the type of writing assessments in Western universities in the disciplines. As Loudermilk (2007) points out, assessments are designed by course instructors, and usually they are the only ones who are able to access the

assignments, apart from the students who wrote them. However this 'occluded' nature of student assignments is slowly being eroded as more corpus data and research findings are disseminated among the EAP community. These studies point to the specific genre types preferred by the disciplines and also how they may be investigated.

2.4 Disciplinary Variation in Academic Writing

This section reviews the literature on disciplinary variation in academic writing to establish the need to investigate genre characteristics in texts belonging to a specific 'discourse community'.

2.4.1 Studies investigating disciplinary variation in Research Articles

The studies reviewed in this section base their analysis of disciplinary variation on the Research Article (RA) genre. These studies are reviewed here since most of the research on disciplinary variation has been conducted on RAs. They are relevant to our discussion of student academic writing because the findings of these studies may also point to disciplinary variation in student writing. This also highlights the lack of research on pedagogic genres and the problems encountered when comparing RAs with academic genres which have other purposes, audiences, rhetorical structures and lexico-grammatical features.

As noted in Section 2.2, although the CARS model is an attempt to identify common schematic structures in RA introductions across disciplines, many of the studies detailed below have used Swales' methods to analyse other sections of RAs.

The textual features of the RA reflect its social contexts (Myers 1985, North 2005) such that 'each text seems to be making a different kind of move in a different kind of game' (Bazerman 1981: 378). Bazerman is here pointing out that the text cannot be taken out of its context and its features viewed as 'isolated conventions' (4). Bazerman's view that genres, in this case RAs, have a recognisable schema but at the same time do not follow a prescriptive formula is acknowledged by the studies reviewed in this section.

Some studies have demonstrated disciplinary variation by comparing the move structure of research article introductions to the structure outlined in the CARS model. For

example, Anthony (1999) analysed 12 Software Engineering (SE) RA introductions to report that although they fitted the broad framework of the CARS model, variations can be found in the steps used to realise the three moves in the framework. He found that the framework does not account for the following steps in SE introductions: an extensive background review, evaluation of research, definitions and examples. He attributes this non-alignment to variability in disciplinary discourses which cannot be captured by a generalised framework.

Several other studies have investigated variation in textual conventions across disciplines by comparing the RAs of two or more disciplines. Some have focussed on particular sections of the RA (Introduction, Methods, Results and Discussion), while others have focussed on the linguistic features and schema of the entire RA.

Basturkmen (2012) compared the move structure of discussion sections from ten Dentistry RAs with ten Applied Linguistics (AL) RAs. The AL framework was found to be broadly applicable to the Dentistry discussion sections. However, disciplinary variation manifested itself in the steps and sub-steps used to realize a move. For instance, the 'explaining findings' step was found in 90% of commenting moves in AL discussion sections while it was only found in just over 50% of Dentistry discussion sections. Construction of complex argumentation using alternative explanations was found in both disciplines, although more frequently in the AL texts. Thus, even though the same patterns of argumentation were used in the two different disciplines, the proportion in which these were used varied. These findings can be seen in terms of the differing epistemologies of the Social and Physical Sciences. The advantage of this in-depth approach to genre analysis is its ability to tease out fine-grained genre variation which includes not only analysing the broad moves in a section but also the steps and sub-steps involved in realising a communicative function. Basturkmen's findings are indicative, but her dataset was too small to allow for any firm conclusions about disciplinary differences.

Studies have also reported variation in the traditional rhetorical sections (Introduction-Methodology-Results-Discussion) of RAs among disciplines. One such variation was noted by Yang and Allison (2003), who found that a supplementary section

labelled ‘Pedagogical Implications’ was frequently included in the Applied Linguistics RAs that they studied. This shows how the nature of a discipline may cause a deviation from the general pattern; the broad aim of applied disciplines-to explore practical applications-influenced the inclusion of a section on pedagogy. Another example of a single-discipline study is Posteguillo’s (1999) study of 40 articles from Computer Science, which found that none of the articles examined adhered to the standard ‘Introduction-Method-Results-Discussion’ format. He ascribes this finding to the relative ‘recentness’ of this emerging discipline which had yet to establish a standard structure.

This type of disciplinary variation in the structural patterning of RAs was also investigated by Lin and Evans (2012) who examined 433 empirical research articles across 39 disciplines in the fields of engineering, applied sciences, social sciences and humanities. Their study was much broader in scope in terms of the size of the corpus and the number of disciplines compared, and found that although many articles followed the IMRD pattern, there were ‘discernible disciplinary variations’ among them such as the inclusion of a separate Literature Review section or a merged Results and Discussion section (Lin and Evans: 2012: 154). Patterns of organisation depended on the nature and orientation of the discipline. For example, the ILM[RD]C (Introduction-Literature Review-Methods-merged Results and Discussion-Conclusion) pattern was the most prevalent in the engineering and social sciences disciplines but was rare in the Applied Sciences and Humanities disciplines. This indicates that there are genre differences across the disciplines even though the genre (RA) is the same. This might possibly be the case with other genres which may be realised in different ways in different disciplines.

The representative studies reviewed in this section clearly demonstrate that there is disciplinary variation in the structuring of Research Articles. There is, however, a dearth of research on pedagogic genres which needs to be urgently addressed because of the pedagogical implications involved.

2.4.2 Sub-disciplinary Variation in RAs

Research has suggested that there are variations in writing genres not only among different disciplines but also among sub-disciplines and related disciplines.

Samraj (2002), for example, found that the 'moves' and 'steps' used to realise communicative functions differed in the RA introductions of two related fields: Wildlife Behaviour and Conservation Biology. Ozturk (2007) investigated the introductions of RAs in two sub-disciplines of applied linguistics: second language acquisition (SLA) and second language writing research (SLWR). She was able to detect sub-disciplinary variation based on Swales' (1990) 'move' analysis framework. The study found that SLA introductions adhered to the CARS model while SLWR introductions did not, the interpretation being that SLWR Research Articles show a deviant pattern because of the status of SLWR as an 'emerging' field, and its interdisciplinary nature. Kanoksilapatham (2012) investigated variation in 12 experimental RAs from each of three sub-disciplines of engineering: Civil Engineering, Software Engineering and Biomedical Engineering. She reported statistically significant variations in the textual organization of these RAs. Her 2015 study on the same dataset revealed that the move structures were identical across the three but the steps used to realise these moves varied. She found, for instance, that the step 'claiming centrality' in Move 1 did not feature in Civil Engineering (CE), something which she attributes to the maturity of this sub-discipline. CE is a discipline which is already well-established (Lin 2014, Posteguillo 1999) and acknowledged by readers as important, while the other disciplines are yet to reach that status and might need to include this move to convince readers of their 'centrality'.

Both Ozturk (2007) and Kanoksilapatham (2015) attribute disciplinary variation to the emerging nature of some disciplines. The term 'emerging' has been defined by Samraj (2005: 144) as referring to 'fields of inquiry that borrow from more than one parent discipline'. This complements Hyland's (1999: 352) definition of established disciplines as 'inhabiting a relatively discrete and clearly identifiable area of study', with research 'proceeding along a well-defined path'.

To sum up, in the studies discussed in this section various factors have emerged as influences on disciplinary variation. These include differing epistemologies, the theoretical or applied nature of the field, the maturity of the discipline or its interdisciplinary (therefore 'emerging') status, and the evolving nature of genres. Section 2.4.3 examines studies that have explored disciplinary variation in pedagogic genres, an area which has received little research attention compared to RAs. This type of research is urgently required because of its pedagogic value. The findings from studies based on RAs cannot be applied to genres that students produce due to their differing purpose, audience, staging and linguistic features.

2.4.3 Studies exploring variation across disciplines in written genres other than RAs

Although most studies of disciplinary variation have been conducted on the RA genre, some researchers have analysed the macrostructure of other genres, especially pedagogic genres. As researchers such as Parkinson (2017) claim, the features of pedagogical genres such as Lab Reports should be studied in their own right since they have a different purpose, audience and context from RAs. Perhaps these differences have not been considered by researchers such as Chen and Baker (2010), Hyland (2008) and Moskovitz and Kellogg (2005) who have recommended that RAs should be used as a model to teach students about disciplinary conventions.

North (2005) compared the essay writing performance of students from science and arts backgrounds in a history of science course and found that the arts students achieved significantly higher scores. She attributes this finding to the variation in disciplinary culture and epistemology which finds expression in discourse styles that are unique to each of these disciplines. North's findings accord with those of other researchers (see Section 2.4) regarding the variability of disciplinary discourse especially between the hard and soft disciplines. The arts students, for example, cited more and made their stance apparent while the science students made no attempt to refer to sources. North claims that this was due to their disciplinary habit of treating statements as accepted facts. North's findings may also bring into question the adaptability of students who are already socialised into a disciplinary culture to the writing conventions of other disciplines. Her findings are in accord with Nesi

and Gardner (2006) who discuss the difficulty medical students face in completing reflective writing assignments, a genre drawing on the conventions of an unfamiliar genre.

Findings related to disciplinary differences in student writing have significant pedagogical implications especially considering the increasingly modular nature of courses. They call into question the effectiveness of teaching general academic writing skills, which is the general practice to support the academic writing of students in many higher education institutions.

The findings of studies such as North (2005) suggest the potential benefit of compiling a substantial corpus of texts belonging to a single genre and discipline for the purposes of analysing textual features and thereby informing pedagogical practice.

2.4.4 Preferred disciplinary genres

Continuing the theme of discipline-specificity, this section reports on investigations into the genres favoured by particular discourse communities with emphasis on the work of Nesi and Gardner (2012).

Nesi and Gardner (2006) explored the similarities and differences in the assessment criteria of practitioners from the sciences and the humanities. For this purpose, they conducted 55 interviews in three UK universities. With regard to similarities, they found that disciplines as diverse as Biology and Psychology frequently assigned classical Research Article writing tasks. Differences in the genres favoured by specific disciplines were also noted. They found that lecturers assigned professional genres in certain disciplines, such as Case Reports for medical students and Site Investigation Reports for engineers. The Essay was the primary assignment for the humanities and the Lab Report as the core assignment of the sciences. These genres, preliminarily dealt with by Nesi and Gardner (2006), were later discussed in more detail in Nesi and Gardner (2012).

The BAWE corpus was the basis on which Nesi and Gardner (2012) identified the genres favoured by specific disciplines. As discussed in Section 2.3.4, they used genre analysis and contextual cues to identify the genres found in the 30 disciplines represented in

the corpus. For example, they report that Case Studies are more prevalent in the Business and Health disciplines while Design Specifications are found mostly in the Physical Sciences in the Computing and Engineering disciplines. The 'Explanation' genre is most frequently assigned in Biology, Chemistry and Engineering. They also found that Engineering assignments were spread across all 13 genre families. Findings of these studies which investigated the preferred genres of specific disciplines have great pedagogical import as they can inform the design of discipline and genre-specific syllabi and learning materials. Further research can build on these findings by investigating the preferred genres in specific disciplines and attempting in-depth linguistic descriptions of these texts which can help students produce successful assignments.

Disciplinary variation can also be seen in terms of the sequencing of genres across university levels. In Arts and Humanities, the proportion of assignments which are Essays decreases significantly from Levels 1 through 4 (91% to 61%) while the proportion of Critiques increases from 4% to 20%. Methodology Recounts are frequently assigned in Levels 1 and 2 in the Physical Sciences while in Level 3 a variety of genres including Reflective Writing are assigned as coursework. Nesi and Gardner report that Engineering students tend to write Product Evaluations, Methodology Recounts and Design Specifications over Levels 1 and 2, and then in Level 3 are assigned a Project which contains all these components as well as a literature review. These findings show that some genres are preferred by specific disciplines at specific levels of study, thus demonstrating the usefulness of discipline-specific genre studies. The detailed characterisation of genres found in such discipline-specific corpora can be aided by linguistic descriptions based on corpus analysis techniques.

2.4.5. Corpus techniques to investigate variation in assignments across student genres and disciplines

Developments in corpus research and technology, and the increasing realization of the futility of 'separating rhetorical process expertise from domain content expertise' (Artemeva 2008) have led to studies of disciplinary and generic variation using the powerful techniques of

corpus analysis. As Sinclair (1991: 100) claims about these techniques, '[t]he language looks different when you look at a lot of it at once'. Some important studies on this sort of variation using corpus techniques are evaluated in this section.

Gardner (2012) traces differences in the presentation of arguments across disciplines by examining the semantic categories of key verbs and adverbs (e.g. *concede*, *admit* to express concession in Philosophy essays), collocations (e.g. *legally impose*) and specific pronoun searches (e.g. the use of *I* in argumentation in English and Philosophy essays). She examined student Essays across five disciplines: Classics, English, Law, Philosophy and Sociology. The study revealed significant variation in the linguistic resources used for argumentation in essays across these disciplines. Nesi (2014) also demonstrates the application of Corpus Query Language (CQL) to the BAWE corpus to examine citation practices of students across the disciplines. She finds large 'quantitative and distributional differences' (Nesi, 2014: 104) in the use of integral and non-integral citations, reporting verbs and citation styles. These studies demonstrate how linguistic features can be examined using corpus techniques and compared across disciplines and genres. As I stated in Sections 2.3.2 and 2.3.3, thorough linguistic descriptions of genres in specific disciplines are a necessary part of uncovering the types of assignments in a discipline. The discussion of corpus analysis techniques in this section suggests that these techniques can be useful to characterise and compare genres both within and across disciplines in terms of linguistic features such as phraseology and other tendencies of academic writing identified in the literature such as nominal features and prevalent frames (e.g. *the N of (the) N).*

These two studies add to the growing literature on using corpus tools to uncover disciplinary and generic variation. For example, there are now more than 100 studies of this type which draw on the BAWE corpus.

2.4.6 Studies using multi-dimensional analysis to investigate variation in academic written genres

Multidimensional Analyses (MDA) are increasingly being used to uncover disciplinary variation in academic and professional written genres (Biber 2006, Gardner, Nesi and Biber

2018, Gray 2011, Nesi and Gardner 2012, and Reppen 2001). The MDA approach was developed originally to identify the differences between varieties of spoken and written English (Biber 1988) and involve identifying clusters of linguistic devices used to realise the communicative purpose of specific registers. MDA is based on a multi-variate statistical test which can reveal the linguistic features that are more or less likely to co-occur in a corpus of texts. Since these co-occurrences are not one-dimensional, multiple dimension continua are then extracted with lexico-grammatical features clustering on the 'positive' and 'negative' ends of these dimensions. Groups of texts can then be placed along each of the dimensions to reveal the lexico-grammatical features that co-occur the most or the least frequently in these texts. For instance, in the first dimension identified in Gardner, Nesi and Biber (2018), 'Compressed Procedural Information' versus 'Stance towards the work of others', linguistic features such as action verbs, pre-modifying nouns and concrete nouns cluster at the 'positive' end while stance adverbials and proper nouns are distributed at the 'negative' end of the dimension. It is thus possible to see that texts that are distributed towards the positive end of this dimension continuum realise grammatical density using a particular set of linguistic features.

MDA studies are based on the assumption that there are systematic differences in the use of the lexico-grammatical features in different registers because individuals make specific choices that are appropriate to the register they are engaging in (Hardy and Römer 2013). Biber et al. (1999: 5) define 'register' as 'situationally defined varieties' of language and consider the texts in their corpus as belonging to different registers ('Conversation', 'Fiction', 'Newspaper Language', and 'Academic Prose'). Biber and Conrad (2009) define register in terms of lexico-grammatical features, contextual factors and communicative purpose. MDA enables the co-occurring features in groups of texts or individual texts to be analysed and mapped in terms of situational variables such as discipline, genre and level of study (Gardner, Nesi and Biber 2018).

Hardy and Römer (2013) used Biber's multi-dimensional analysis (MDA) approach to explore disciplinary variation as manifested in the MICUSP corpus. They analysed the co-occurring linguistic features of texts to plot them on the four dimensions chosen for their

study. For instance, their first dimension, ranging from 'Involved, Academic Narrative' to 'Descriptive, Informational Discourse' was characterised by high use of verbs, complement clauses, personal pronouns, and past tense forms at the positive end, and relatively frequent incidences of nouns, nominalizations and adjectives at the negative end. Genres such as argumentative essays and creative writing from the Arts and Humanities and Social Sciences disciplines were mostly placed at the 'Involved' end of the dimension while genres such as 'Reports' and 'Research Papers' from the Physical and Life Sciences disciplines were placed at the 'Descriptive, Informational' end.

Hardy and Römer's (2013) findings can be compared to Gardner, Nesi and Biber's (2018) MDA study. This recent study revealed that the hard and soft disciplines employ different sets of co-occurring lexico-grammatical features to achieve density in writing. The main contribution of this study is the discovery of two types of density in academic writing: the first type, favoured by the 'hard' sciences, is characterised by pre-modifying nouns, common nouns, passives, action verbs, quantity nouns and concrete nouns, while the second type of density, found more frequently in assignments belonging to the Social Sciences is realised by nominalisations, attributive adjectives, long words and abstract nouns. Nesi, Gardner and Biber's (2018) findings can be explored further by tracing the two types of density across genres in a discipline or across genres in different disciplines. Such exploration could take a corpus-based approach starting from the linguistic features in question, to reveal interesting variations in the way genres and disciplines realise density.

The studies reviewed in Sections 2.4.4 and 2.4.5 demonstrate the potential of corpus and MDA analyses to reveal linguistic characteristics of texts and enable comparisons of individual and sets of texts across the situational variables of genre, discipline and levels of study. They establish that the linguistic features of genres co-occur and that these clusters of features differ across these variables. The findings from MDA studies could be followed up by investigations of some of the clustering linguistic features in genres in specific disciplines. For example, genre and discipline specific corpora could be used to investigate whether the texts in pedagogical genres in a single discipline use the linguistic features that cluster at the positive end of Gardner, Nesi and Biber's (2018) Dimension 1, (Compressed Procedural

Information), or whether they realise grammatical density using other linguistic devices which are clustered around the positive end of Dimension 4 ('Information Density')

2.5. Teaching writing to engineers

This section will focus on studies of written engineering texts and their pedagogical implications. This is still a somewhat under-researched area, although there have been some influential studies of engineering genres.

For example, Gardner (2008) compared 250 Engineering texts with 60 History assignments in the BAWE corpus, and found that Engineering students wrote a variety of genres, distributed differently across the years of study. She found that Exercises and Explanations, which are more likely simply to report on the current state of knowledge in the field, gave way to more original and evaluative genres such as Design Specifications and Critiques at the higher levels of study. Gardner concluded that Engineering students have greater demands placed on them in terms of writing than History students, because of the greater range of genres they need to be familiar with and the wider range of language features required to realize them.

Other studies have attempted to study specific linguistic features characterising Engineering writing. Leedham (2011) investigated the similarities and differences between Chinese and English students' undergraduate assignments in three disciplines (including Engineering), in terms of key words, connectors, pronouns, visuals and lists, and text formatting. She concludes that even though there are differences between Chinese and English L1 students in their preferences for certain lexico-grammatical and textual features for realising the same communicative purpose, both groups are nevertheless considered proficient writers and therefore, the notion of what constitutes 'good' writing should be expanded and made more flexible to accommodate a wider range of writing styles.

Leedham and Fernandez-Parra (2017) studied the difference in the usage of first person pronouns among Chinese, Greek, and English L1 Engineering students by examining assignments submitted to British universities. The Greek students used *we* twice as much as the Chinese students, while the British students used it the least. Regarding the singular first

person pronoun *I*, the Chinese used it the least while the English and Greek L1 students were similar in the frequency of their use of this pronoun. Each group also showed preferences in the functions for which these pronouns were employed. For example the 'Representative' use of *we* was preferred by the Chinese students; *we* was used to denote the student group working on the assignment, or to a larger group such as engineers or manufacturers. Leedham and Fernandez-Parra conclude that, in the Engineering assignments that they examined, students' writing styles are influenced by their language, culture and values, and that the use of personal pronouns is acceptable in the Engineering disciplines.

A noteworthy set of 'genre-studies' in Engineering are those conducted by Conrad and her co-authors (2014, 2016, 2017, 2018) in the area of Civil Engineering (CE) writing. These have combined MDA, genre analysis, corpus linguistic techniques, and interview data to analyse and compare practitioner writing with student writing, and to demonstrate the areas where students can emulate professionals. Based on the results of this comparison, the move structure of genres and the use of certain linguistic features such as short sentences without relative clauses and active voice is recommended for students who engage in genres that simulate workplace writing such as 'Bridge Descriptions' and 'Reports'. Conrad et al. (2016) describe the implementation of a pedagogical approach that involved close collaboration with Civil Engineers, focusing on the problems that students face in writing, the development of course material to address these problems and the assessment of the effectiveness of the material based on practitioner feedback. The work of Conrad and her colleagues contributes to the understanding of the linguistic requirements of those student genres in CE that simulate workplace genres (e.g. 'technical memos') and leads the way for successful collaboration between subject teachers and practitioners for the implementation of relevant and effective teaching strategies.

However, Conrad (2014, 2017) and Conrad et al. (2016, 2018) did not focus on the pedagogical genres in CE which students may never actually produce in their professional lives but which nevertheless are important for success at the university.

Other studies with similar aims to those of Conrad (2017) and Conrad et al. (2016, 2018) report successful pedagogical interventions to raise genre-consciousness among students, but have been quite small-scale in scope, with some studies even reporting on the performance of a single learner or a few learners (Artemeva 2008, 2009, Cheng 2007). These studies claim that students were able to transfer the genre writing skills they acquired from their EAP course to their workplace requirements as a result of their new-found genre awareness. However, these studies do not account for other factors that may have contributed to the success of these students, such as the assignments set by the subject lecturers, which may have prepared the students for workplace genres, or other regulatory mechanisms such as professional bodies or the industry affiliations of the institutions concerned, whose feedback could have influenced academic departments to design their assignments to simulate certain workplace genres. Although research on coursework which simulates workplace genres is important, other pedagogic genres deserve separate investigation because students need to be able to write successful assignments in order to graduate. Therefore, studies can build on the insights provided by Conrad et al., who focussed on only workplace simulating genres, by identifying all the genres assigned in Civil Engineering to obtain a more in-depth linguistic characterisation of pedagogical genres.

Some pedagogical resources are available for Engineering students, such as the British Council's *Writing for a Purpose* website which contains materials and activities on different types of writing tasks for different disciplines. In addition Durrant (2014), Mudraya (2006), and Ward (2009) have compiled and evaluated word lists for engineers, and articles such as Wolfe, Britt and Alexander (2011), and Gardner (2016) discuss pedagogical interventions designed to teach Engineering writing.

Despite the work done by researchers on Engineering writing, such as attempts to describe the genres in Engineering in BAWE (Nesi and Gardner 2012) and Conrad's (2017, 2018) work in Civil Engineering, there still remains a gap in our knowledge of the student genres assessed in specific Engineering disciplines. For instance, studies intended to identify student genres in Civil Engineering based on a carefully designed discipline-specific corpus have not been conducted. These types of investigations are a pre-requisite for implementing

pedagogies designed to facilitate the acquisition of genre-awareness and lexico-grammatical features of specific engineering genres.

2.5.1. Engineering Vocabulary

Just like the studies in 2.5 which attempted to raise awareness in engineering students about their disciplinary genres, other studies have been undertaken to identify engineering vocabulary in an attempt to provide more discipline-specific support to students. These studies recognise that familiarity with the relevant disciplinary vocabulary is essential for communicative success (Durrant 2014, Nation and Hunston 2013).

Many prominent researchers (Coxhead 2000, Nation 2001) have accepted the three-way categorisation of vocabulary: 1) general purpose words found in a range of texts, both academic and non-academic; 2) 'academic vocabulary' (Nation 2001) or 'sub-technical vocabulary' (Cowan 1974) which comprise the 'shared vocabulary of several fields of study'; and 3) 'specialized vocabulary of one particular field or part of that field' (Nation and Hunston 2013: 187). As Ward (2009) points out, there have been claims that the teaching of 'technical vocabulary' is the responsibility of subject lecturers while EAP teachers can support the teaching of 'academic vocabulary' because these words are accessible to them. Therefore, many compilations of academic vocabulary lists have been attempted with a view to supporting students in their academic writing requirements.

However, such initiatives can be said to have met with only partial success due to several factors, including long word lists which are difficult for non-native students to learn, the difficulty of arriving at a list of 'general' academic words that are equally frequent or useful across disciplines and genres, the neglect of technical words, which might not be taught by the subject lecturers (Ward 2009), and the difficulty of separating the three types of words, as items in the 'academic' list or 'general' list might cross over to the 'specialized' vocabulary list when they take on a specialized meaning in particular disciplines (e.g. *stress* in Civil Engineering). With regard to academic words, for a word to be truly learned, students also need to be familiar with their grammatical behaviour, phraseology, collocation and other

characteristics (Nation 2015), which also explains why just a list of words may be of little pedagogical value.

Some studies have attempted to compile more discipline-specific lists for broad disciplinary areas such as Engineering (Hsu 2014, Mudraya 2006). Some of the arguments against an academic word list also apply to these engineering word lists. Mudraya's corpus, for example, contained texts from nine Engineering disciplines from Electrical Engineering to Computer Programming. There are considerable differences among these sub-disciplines and it would be impossible to arrive at a 'high-coverage' list whose items are shared by these diverse disciplines. In this regard, it is interesting to note that Hsu (2014), who also compiled an Engineering word list based on 20 Engineering sub-disciplines, reports variations in the number and type of words students need to be familiar with in the different sub-disciplines. For instance, Civil and Mechanical Engineering were the least demanding in terms of lexis, while Marine and Biochemical Engineering students needed to acquire more than double the number of discipline-specific words.

These studies underscore the necessity of identifying vocabulary that supports the specificity of disciplines, as pointed out by many researchers (e.g. Durrant 2014, Ward 2009). This is especially important in EFL contexts where many students find it taxing to learn a large general academic vocabulary, some of which they might not even encounter in the course of their academic reading and writing.

This section has established the need for studies which aim to identify the vocabulary used in specific sub-branches of Engineering. Ways need to be devised of exposing learners to these words in context so that they can be acquired on the basis of their grammatical behaviour, collocations and other characteristics.

2.6. Studies of academic writing in the 'Expanding Circle'

Most of the studies conducted in countries which I group as belonging to the category of *English as a lingua franca country in the 'expanding circle'* seem to be dated. For example, Flowerdew (1993) and Mustafa and Swales (1984) undertook research in disciplinary writing decades ago. More recently, some studies (Crompton 2011, Randall and Groom 2009) have

been conducted on student writing in the Arabic context; however, these studies are primarily based on learner corpora and deal with general grammar and spelling difficulties that students encounter. Learner corpora have been defined as 'corpora containing data produced by L2 learners…both foreign and second language learners' (Gilquin, Granger, and Paquot 2007). These corpora, such as the *International Corpus of Learner English* (Granger et al. 2009), mainly contain essays on general topics and are used to examine and correct syntactic, lexical, semantic and pragmatic errors produced by such learners. Such texts have little to do with the type of writing students are required to produce in their disciplinary courses, so the discourse organization and lexico-grammatical features of disciplinary genres cannot be explored. The findings of studies based on such corpora are less likely to be useful for students enrolled on undergraduate or graduate courses in disciplines such as Engineering.

A few studies (El Malik and Nesi 2008, Al Shahrani 2015) have analysed the writing of Arab L1 writers to investigate the similarities and differences in their use of certain linguistic features. El Malik and Nesi (2008) examined 20 medical RAs published in prestigious medical journals to investigate the differences between Sudanese and British writers, and Al Shahrani (2015) compared the use of metadiscourse in the discussion and conclusion sections of doctoral dissertations between native English and L1 Arab writers. Both these studies report differences in the writing of native and non-native speakers. El Malik and Nesi noted that British writers tended to hedge more to achieve an objective author stance and used more nominalisations to make their writing more succinct. However, it should be noted that the groups of L2 writers in both studies were successful writers as their articles and dissertations were published. These findings suggest that although they belong to our 'expanding circle lingua franca' group, the Arab writers in El Malik and Nesi (2008) and Al Shahrani (2015), possessed an adequate syntactic and lexical repertoire to gain access to their discourse communities. Like these writers, I argue that university students in countries such as Oman or Dubai require academic writing skills that are 'adequate'. Writers in these contexts need to be able to communicate clearly and accurately; elegance in academic writing does not seem to be a priority in these contexts. Therefore, the writing

support that is offered to such students can be accordingly tailored to be 'fit for purpose' in their contexts.

This section has established the need for more studies of academic writing in the 'expanding circle'. It has argued that academic writing skills that are 'fit for purpose' are what is needed in these contexts. It follows that texts collected from proficient student writers in institutions located in these contexts are better exemplars for students, teachers and researchers, and that it is these texts that should be used to create corpora for research. It is important to consider what types of research can be carried out using such corpora, one important area being discipline-specific phraseology.

2.7. Phraseology and Lexical Bundles

The phraseological nature of language continues to intrigue researchers and influence many domains of enquiry, including Corpus Linguistics, Grammar and Psycholinguistics (Wray 2012). Researchers such as Firth (1957), Bolinger (1976), Pawley and Syder (1983) and Sinclair (1991) have emphasized the importance of phraseology in describing the lexicon of English by pointing out the prevalence of prefabricated units in spoken and written discourse. Much of the early research on word combinations was theoretical in nature because of the lack of empirical methods to retrieve word-sequences from large amounts of data. For example, Pawley and Syder (1983), in their seminal paper on recurring word sequences in speech, discuss the role of such sequences in making language fluent and natural, their ability to increase processing speed, and the challenges that second language learners face in mastering their use.

Sinclair (1991) used corpus linguistic methods to illustrate his idiom principle, which emphasizes the inseparability of lexis and grammar in language. This principle posits that there are a number of 'semi pre-constructed phrases' (1991: 110) at the disposal of language users for use in specific contexts, and that these phrases may not make sense when the words are analysed separately (e.g. phrases such as *of course* should be analysed as wholes rather than according to individual words). Pawley and Syder (1983) and Sinclair (1991) focus on the availability of a stock of ready phrases (fixed or variable to a degree) which the user can

retrieve as a whole to be used in the appropriate context. The implication that these words are structurally and semantically complete is indicated by the use of the word 'phrase'.

Phraseological restrictions on the linguistic behaviour of words have been the focus of much work in this area (Goldberg 1995, Hoey 2005, Hunston and Francis 2000). Frameworks such as Construction Grammar (Goldberg 1995) and Pattern Grammar (Hunston and Francis 2000) associate grammatical structures and specific semantic sets. Hoey's (2005) work on Lexical Priming explains how words are primed for use by encounters with them in differing situations and genres. This links to the work of Wray (e.g. 2002), who has developed a model for formulaic sequences as used by first- and second-language learners. This work brings a psycholinguistic focus to the study of formulaic language by pointing out that one of its main functions is to decrease processing load (Wray and Perkins 2002).

Other work in phraseology has sought to find automatic and semi-automatic ways of retrieving phrases from corpora. Renouf and Sinclair's investigations of productive and pre-selected discontinuous sequences of words (1991) is a notable contribution in this respect. They term these sequences of two or more closed class words which frame a variable slot, 'collocation frameworks'. For example, in *a (n) * of*, the asterisk indicates the presence of one intervening item. Their work and later studies (Eeg-Olofson & Altenberg 1994, Marco 2000) have examined the degree of variability of the intervening items in these 'collocational frameworks' and the type of words, function or content, which fill the variable and fixed slots.

An approach which represents a methodological development from collocational frameworks is the *phrase-frame* of frequently occurring discontinuous sequences (Biber 2009, Vincent 2013). Phrase-frames (p-frames) are automatically retrieved fixed strings of items in which one of the items can vary. This method has been used to explore phraseological variation and associate particular frames with specific genres and registers. For example, Biber (2009) pointed out that the phraseology of writing and speech differs in that the former has frames with internal variable slots that are filled with content words, while the latter was characterised by fixed multi-word sequences. Studies have also discussed

the most likely words or semantic word classes that are likely to occur in the variable slots in particular frames. Römer (2010) investigated the types of words occurring in the variable slots in the frames, *it * be interesting to* and *it would be * to* and found that phraseological restrictions applied, in that, the former frame was only completed by the modals *will*, *would* and *might*, while the intervening word in the latter was usually an evaluative adjective. Nevertheless, work in phrase-frames is still at a relatively early stage and criteria for selecting specific p-frames to study are still under-developed, which makes them a less attractive proposition than the more developed work on *n-grams*, automatically retrieved recurrent fixed strings of words.

Other work in phraseology has also drawn attention to the association between the frequency of particular word sequences and the genres in which they are most prevalent (Ädel and Erman 2012, Ebeling and Hasselgård 2015, Hyland 2012); such work has clear pedagogical implications. Studies of this type emphasize the role of formulaic language in discourse community membership. As Wray (2006: 593) points out, 'when we speak, we select particular turns of phrase that we perceive to be associated with certain values, styles and groups'. A particularly influential data-driven approach to investigating the association between genres and word sequences is the automatic retrieval of frequent fixed word sequences from corpora. These recurrent multi-word sequences are generally known as either *n-grams* or *lexical bundles* (Biber et al. 1999). The main characteristics of Lexical Bundles (LBs) are that they are non-pre-defined, recurrent, fixed, and are not typically perceptually salient.

This frequency-driven approach to the study of multi-word sequences has been adopted by numerous studies, yielding convincing evidence that LBs form part of an interesting linguistic phenomenon which can help differentiate different types of text (Altenberg 1998, Biber et al. 1999, Fletcher 2012, Hyland 2008, 2012).

In the field of EAP, the prevalence and distribution of repeated word sequences in specific registers and genres have been explored in studies such as Biber et al. (2006), Cortes (2004) and Hyland (2008). These studies have indicated that specific disciplines and genres favour different sets of word sequences. It follows from this finding that sensitizing learners

to the use of appropriate phraseological expressions should ease their entry into their discourse communities. The need for second language learners to master recurring multi-word sequences in specific genres in order to support their academic endeavours has been summed up by Ellis, Simpson-Vlach and Maynard (2008: 377):

> *Every genre has a characteristic form of expression, and learning to be effective in the genre involves mastering this phraseology.*

Hyland (2008: 5) also emphasizes the importance of learners being familiar with these multi-word sequences in order to be conversant with the language of their disciplines.

> *Gaining control of a new language or register requires a sensitivity to expert users' preferences for certain **sequences of words** over others that might seem equally possible. So, if learning to use the more frequent fixed **phrases** of a discipline can contribute to gaining communicative competence in a field of study, there are advantages to identifying these clusters to better help learners acquire the specific rhetorical practices of their communities.*

However, the terminology adopted by writers in referring to the study of these repeated word sequences is somewhat fuzzy, as is demonstrated in the quotation above, which equates 'sequences of words' with 'phrases' (Hyland 2008: 5). Not all sequences are 'phrases' with structural completeness, although some may be, such as *on the other hand*. In this review, recurrent word sequences generated from corpora by computer software will not be referred to as 'phrases', differing in this respect from some other studies (e.g. Chen & Baker 2010, Hyland 2008).

A review of the corpus linguistic literature on recurring word strings reveals in fact that there is a confusing array of labels to refer to highly similar phenomena, including: *clusters* (Scott 1997); *n-gram*s (e.g. Stubbs 2007); *recurring word combinations* (Altenberg 1998, De Cock 1998) and *lexical bundles* (Biber et al. 1999). This thesis will adopt the term Lexical Bundles (LBs) to refer to recurring fixed-length multi-word sequences which are retrieved from corpora using specific frequency and distribution criteria, mainly because many studies have adopted this term, as the review below demonstrates.

2.7.1. Identification of Lexical Bundles

Three criteria are typically used to identify LBs in corpora: frequency of occurrence, dispersion across texts in a corpus, and length of the LBs. In terms of frequency, Biber et al. (1999) adopted a cut-off frequency of 10 times per million words for their 9 million-word corpus of Academic Prose and Conversation. Later studies have established more conservative frequency cut-off points for LBs: Pan, Reppen and Biber (2016) set the cut-off frequency at 40 times per million words (pmw) for four-word lexical bundles (they used two sub-corpora of 500,000 words each); Biber and Barbieri (2007) adopted the same cut-off (40 pmw) for a range of sub-corpora of varying sizes (from about 39,000 to 1.2 million words) used in their study; Chen and Baker (2010) established a cut-off frequency of 25 times per million words for their three sub-corpora with an average size of 150,000 words; Cortes (2004) set the figure at 20 pmw for the range of sub-corpora of varying sizes (between about 93,000 words to 1 million words); Hyland (2008) set the frequency at 20 times per million words for his corpus of 3.5 million words.

Setting a lower frequency threshold will result in the retrieval of a greater number of LBs. Biber et al.'s LB studies (1999) set the lowest frequency cut offs perhaps because the primary aims of the studies were to investigate the frequencies and structurally classify the sets of LBs. In this case, the researchers may have considered it useful to generate a greater number of LBs. In other studies (Pan, Reppen & Biber 2016: 63), the frequency thresholds seem to be arbitrarily set with justifications such as 'the standard in most previous lexical bundle studies' for using a 40 pmw cut off. One reason for later studies such as Hyland (2008), Chen and Baker (2010) and Biber and Barbieri (2007) setting higher thresholds could be that they wanted to limit the number of LBs generated in order to identify LBs specific to a register or a discipline. In any case, a review of the literature on frequency thresholds reveals that it leaves researchers with some leeway in terms of cut-off points.

The dispersion criterion for identifying LBs is used to eliminate the possibility of individual writers' preferences for idiosyncratic sequences skewing the results (Pan et al. 2016). Again, with regard to the range of texts, linguists have established very different cut-off points. Dispersion rates are reported in terms of the number of texts, but when these are

converted into percentages the variation is evident. These methodological differences may affect the findings of studies that have compared LBs retrieved using widely differing dispersion criteria. For example, Biber et al. (1999) and Biber et al. (2004) set their dispersion criteria as 5 texts which amounts to 0.14% for their Conversation sub-corpus (3,436 texts) and 1.22% for their Academic Prose sub-corpus (408 texts). The findings of studies comparing LBs from Conversation and Academic Prose retrieved using these dispersion rates may have to be treated with caution due to these arbitrary cut offs, established irrespective of the number of texts in the two sub-corpora.

LBs ranging in length from two to seven words have been studied, though four-word LBs have been identified by many researchers as the ideal length to study (Biber et al. 2004, Hyland 2008), as they are not as frequent as the two or three-word LBs nor as rare as five-word LBs, and they tend to be more complete than the shorter sequences. It has also been noted that when LBs of different lengths are retrieved, many of the shorter LBs are subsumed under the longer ones; this may lead to the same or similar LB being counted many times over. Even when LBs of the same length are retrieved, they may be counted more than once when they overlap. Biber et al.'s (1999: 993) example of this nesting process of smaller LBs in longer ones is given below.

do you want; you want me; want me to; me to do →
do you want me; you want me to; want me to do →
do you want me to; you want me to do →
do you want me to do.

This means that studies of LBs need to be clear about whether or not they have accounted for this potential overlap. This phenomenon may also affect figures quoted for the proportion of a corpus that is made up of LBs, for example, the claim made by Biber et al. (1999) that 21% of their Academic Prose sub-corpus consists of LBs.

2.7.2. LB features and their contribution to discourse analysis

LBs are fixed and cut across syntactic and semantic barriers because they are automatically retrieved rather than being based on pre-defined structurally complete units (Biber et al.

1999). Their meaning and structural characteristics are often only apparent when the co-text is taken into consideration. This has pedagogical implications, especially because it has been noticed that most LBs are incomplete structurally. For example, Biber et al. (1999) found that less than 5% of the LBs in academic prose are complete structural units.

The advantage of the LB analysis method is that multi-word sequences can be retrieved empirically from large amounts of data, providing an overview of their use in a particular language variety. LB analysis is a 'big picture' approach which involves examining the frequencies of LB types and categorizing them into structural and functional categories in specific language varieties. This frequency-led approach is particularly useful in studies that aim to characterize a very large set of texts belonging to one discipline or genre, and can be complemented by qualitative analysis of passages from a sample of the texts under analysis. For example, the co-text of specified LBs can be examined through concordance lines.

The weakness of this approach is that LBs are not always fully analysable out of context. This problem is hard to remedy since it is impossible to examine each instance of use when dealing with thousands of examples. It is also not possible to identify discontinuous frames (e.g. *not only…but also*) using this method (Nesi and Basturkmen 2006) and other variations (e.g. tense, order of elements inside the LB) since LBs are fixed sequences of words. The program developed by Fletcher (2012), *kfNgram*, gets around the problem of discontinuous frames, to some extent, by automatically identifying the most frequent frames in a corpus (e.g. *the N of (the) N*) and detecting common frames by allowing for variation within n-grams. However, this software has only been used in the occasional study (Cunningham 2017).

The findings of some of the first influential studies on LBs, based on the 40-million word *Longman Corpus of Spoken and Written English*, are reported by Biber et al. (1999) in the reference grammar, *Longman Grammar of Spoken and Written English*. In fact, the term 'Lexical Bundle' was coined by Biber and his colleagues, who were among the first researchers to study LBs.

The *Longman Corpus of Spoken and Written English*, representing both British and American language varieties, contains four core registers (Conversation (about 6.4 million words), Fiction (about 5 million words), News (about 5.5 million words), Academic Prose (about 5.3 million words) and two supplementary registers (Non-conversational speech (5.7 million words) and General Prose (about 6.9 million words). Biber et al. retrieved LBs of different lengths (4, 5 and 6-word LBs) from the Academic Prose and Conversation sub-corpora, identified common LB structures, and attempted some 'preliminary' discussion of the discourse functions they were associated with. They found that LBs constituted a sizeable proportion of these two registers (30% in Conversation and 21% in Academic prose) and that there were considerable differences in LB use in terms of frequency and range.

2.7.3. Biber et al.'s (1999) structural framework for analysing LBs

Although LBs do not have a complete structure, Biber et al. claim that they 'fall into several basic structural types' (1999: 991). The LB patterns from the structural framework they developed are shown in Table 2.4, with examples of four-word LBs found in the Hyland corpus. The Hyland corpus comprised research articles, Masters theses, and PhD dissertations written by experts, doctoral candidates and Masters students. Four disciplines were represented in the corpus: Electrical Engineering and Microbiology (Applied and Pure Sciences), and Applied Linguistics and Business (Social Sciences).

As can be seen in Table 2.4, seven specific patterns were found in the Hyland corpus and one 'other' category was added to classify those LBs that did not fit into the structural framework. The main structural patterns of LBs in Academic Prose identified by Biber et al. (1999) were collapsed into three main types in other studies (Chen & Baker 2010, Pan, Reppen & Biber 2016): those that contain noun phrases (NP), preposition phrases (PP), and verb phrases (VP). These studies report that NP and PP bundles are predominant in the academic writing of proficient writers. Biber et al. (2004) also found that 70% of the LBs in Academic Prose are phrasal and mostly consisted of noun phrases (NP) or a sequence that bridges across two prepositional phrases (PP) (e.g. *as a result of*).

However, the expert texts that these studies examined cannot be equated to pedagogical genres and further research is required to see the distinction of bundles across genres and disciplines.

Biber et al. (1999) found that the structural correlates of LBs in their sub-corpora were different from each other. Conversation LBs usually consisted of the pronominal subject followed by an extended verb phrase (e.g. *I thought it was*) while Academic Prose LBs typically constituted parts of noun phrases and prepositional phrases (e.g. *the nature of the, as a result of*). The associations Biber et al. made between the structure and the functions of LBs led to the development of their functional taxonomy. For example, Biber et al. (1999: 1014) found that many LBs with the *noun phrase + of-phrase* fragment structure (*the size of the, the base of the*) fulfil important functions of physical description 'including identification of place, size and amount.'

Biber et al. also report some other characteristics of LBs. They found that four-word LBs in Academic Prose commonly end with function words. They also identified 1000 frames which were 'extremely productive'. These were constituted of nominal and prepositional elements such as 'the…of the', with slots that could be filled by many words (e.g. *end, base*). Frames like these are also discussed by Fletcher (2012), Hunston and Francis (2000), and Sinclair and Renouf (1991). Another feature of LBs that Biber et al. (1999) discuss is the lack of salience of most of the LBs retrieved, due to the fact that they are incomplete structural units.

Hyland (2008) used Biber et al.'s (1999) structural framework to categorise four-word LBs in a corpus of 3.5 million words. Hyland used a frequency cut-off of 20 per million words and a dispersion rate of 10%.

The structural patterns from Biber et al. (1999) found by Hyland (2008) in his corpus are given in Table 2.4, along with the examples he provided.

Table 2. 4. 4-word bundles in the Hyland Corpus (2008) using the framework of Biber et al. (1999: 996)

The structural patterns identified by Hyland (2008) are similar to those in Biber et al.'s (1999) classification except that he found LBs for only seven of the patterns, probably because of the differences in the corpus compositions. Biber et al.'s (1999) corpus contained academic books in addition to research articles while Hyland's corpus consisted of research articles, student dissertations, and doctoral theses. This may be the reason why more patterns were found in Biber et al.'s (1999) corpus. Rezoug and Vincent (2018) also failed to find the additional patterns in their corpus of student Master's dissertations, although some might have been included in the 'other' category. The stricter cut-offs used by Rezoug and Vincent (2018) (40 times pmw) and Hyland (2008) (20 times pmw) may also have accounted for the absence of these patterns.

Researchers have not reported many problems with the structural classification systems of Biber et al. (1999) and Hyland (2008) because the categorizing procedure is fairly

straightforward and does not involve a great deal of interpretation which requires examination of the expanded context of use.

2.7.4. Biber et al.'s functional (2004) framework for analysing LBs

The development of a functional taxonomy for analysing LBs is an important development in LB studies. Biber et al. (1999) drew attention to certain associations between the structure of LBs and the distinct (and not so distinct) discourse functions of LBs (see 2.7.4). LBs belonging to the various functional categories occur in differing proportions across genres and disciplines, as studies in the later sections (see 2.7.4.1) will reveal.

The three primary discourse functions of LBs proposed by Biber et al. (2004) are: *Stance Expressions*, *Discourse Organizers*, and *Referential Expressions* (see Table 2.6). This functional framework reflects Halliday and Matthiessen's (2004) three metafunctions of language: the interpersonal, the textual, and the ideational. Stance bundles show attitude or express degrees of certainty and are divided into five subcategories: 'epistemic' (*the fact that the*), 'desire' (*I don't want you to*), 'obligation' (*you have to do*), 'intention/prediction' (*what we are going to*), and 'ability' (*to be able to*). Discourse organizing bundles show relationships between preceding and subsequent discourse and serve two sub-functions: 'topic introduction' (*What I want to do*) and 'topic elaboration' (*has to do with the*). Referential bundles identify and specify attributes of an entity and are further classified on the basis of three functions: 'imprecision indicators' (*or something like that*), 'attributes' (*the nature of the*), and 'time/place/text reference' (*at the bottom of*). The LBs in a fourth additional functional category, *Special Conversational Expressions*, are mainly used to fulfil the discourse function of expressing politeness, making inquiries, and reporting (e.g. *thank you very much*). Table 2.5 illustrates this functional taxonomy.

Some materials have been removed due to 3rd party copyright. The unabridged version can be viewed in Lancester Library - Coventry University.

Table 2. 5. Functional Classification of LBs (Biber, Conrad and Cortes 2004)

The functional taxonomy developed by Biber, Conrad and Cortes (2004) is an extension of the discussion of LB functions in Biber et al. (1999). Biber et al. (1999) connected the structure and function of LBs by proposing the primary functions performed by a few of the LB structures they identified. For example, they report that most of the bundles in their Conversation sub-corpus having the structure 'Personal pronoun + lexical verb phrase (+complement-clause fragment) (e.g. *I don't know what, I can't remember what*) are used to express stance, or report personal feelings or thoughts (see Biber et al. 1999: 1003-1023).

Biber et al.'s (2004) study intended to propose discourse functions for the common LBs in their Textbook and Classroom Teaching sub-corpora from the 2.7 million-word T2K-SWAL corpus of US university language. They developed the functional framework for LBs (see Table 2.5) by examining their co-text and inductively grouping the LBs with similar functions.

Some studies have used this framework while others have adapted it. The next section will review two studies that have used it and one that has adapted it.

Biber et al. (2004) report three common patterns of LB discourse function distribution in the T2K-SWAL corpus: (1) Stance Bundles are 'extremely common' in both the spoken registers (2) Discourse Organizers are the most common in Classroom Teaching and also moderately common in Conversation (3) Referential bundles are common in Classroom Teaching and Textbooks and are found less frequently in Academic Prose. The highest normalized frequency of LBs across all three discourse functions was found in Classroom Teaching. Biber et al. (2006: 147) explain that this register places high communicative demands on the user because of its affinity to 'involved spoken discourse' which involves the use of Stance bundles, high information density, which involves the use of Referential

bundles, and the need to structure the discourse, which explains the use of Discourse Organizers. The first two main findings (Biber et al. 2004) about Stance and Discourse bundles being common in the spoken registers is hardly surprising considering the need for speakers to organize their utterances for their listeners. The third finding regarding the prevalence of Referential bundles in textbooks and their comparatively lower frequency in Academic Prose reflects the information density of these genres. It also highlights genre variation in the distribution of LBs. Biber (2006) discusses disciplinary differences in the specific ways in which Referential bundles are used. For example, 45% of all the Referential bundles in Engineering textbooks (sub-corpus of T2K-SWAL) express quantity or mathematical meanings (e.g. *the magnitude of the, of the number of*) while 50% of the Referential bundles used in Humanities textbooks are 'intangible' (e.g. *the nature of the, the notion of a*). For studies seeking to describe the phraseology of pedagogic genres in particular disciplines, the need to investigate variation in the ways LB types are used in them is highlighted here. Some studies (Hyland 2008 and Rezoug and Vincent 2018) have followed this line of inquiry by investigating disciplinary variation in the distribution of LBs by function and the various functional purposes they are used for across the disciplines.

However, application difficulties affect the reliability of this framework too. Researchers investigating student writing or other genres not represented in the T2K-SWAL corpus may find it hard to use this taxonomy as there is not enough explanation of the findings. Additional problems are that many sub-categories are not clearly defined and are often left open to interpretation. Thus studies of LBs should perhaps be more cautious regarding their conclusions and the pedagogical implications that they propose. Although Biber (2006: 172) concludes that LBs are 'readily interpretable in both structural and functional terms', in terms of function especially, it would be difficult to concede this claim.

2.7.5. Hyland's functional framework for analysing LBs (2008)

Hyland (2008) adapted Biber et al.'s functional framework to analyse LBs in the 3.5 million Hyland corpus (the same as the one used for developing his structural framework) consisting of texts contributed by experts, doctoral candidates, and Masters students. The terms adopted by Hyland to examine the discourse functions of the LBs in his corpus are 'Research-

Oriented' (RO), 'Text-Oriented' (TO) and 'Participant-Oriented' (PO) (see Table 2.6). According to Hyland's scheme (2008: 13-14), Research-Oriented bundles include those that help the writer express 'activities and experiences of the real world', Text Oriented bundles help in text organization and 'its meaning as a message or argument' , and Participant Oriented bundles 'are focused on the writer or reader of the text' .

Table 2. 6. Functional Classification of LBs (Hyland 2008)

In effect, Hyland (2008) renamed Biber et al.'s (2004) original categories, and re-arranged and left out some of the sub-categories, although some similarities can still be seen. Most of the sub-functions of Hyland's 'Research Oriented' bundles can be equated to the sub-functions of Biber et al.'s 'Referential Expressions' (e.g. Biber et al.'s 'place' and 'time' sub-categories (1999) can be equated to Hyland's 'Location' sub-category, and the former's 'Quantity Specification' category is the same as Hyland's 'Quantification' sub-category). The additional sub-functions under the Referential Expressions category in Biber et al. (2004) include 'Identification Focus' (e.g. *of the things that*, *that's one of the*) and 'Imprecision' (e.g. *or something like that*), both of which contain bundles which are mainly confined to spoken registers, which may explain their omission from Hyland's framework.

Hyland (2008) explains that his Participant Oriented category provides a framework for analysing two main types of meaning in a text: 1) the stance that the producer of the text holds and 2) the way the writer engages with the reader through the text. In terms of Hyland's 'Participant Oriented' category, which equates to Biber et al.'s (2004) Stance Expressions, the differences seem quite large, since Hyland's (2008) framework contains only two sub-categories ('stance' and 'engagement') rather than the original five. However, four of the sub-categories in Biber et al.'s 'Stance Expressions' category (Prediction, Obligation, Epistemic, Ability) are collapsed under one main category 'Stance' in Hyland (2008). Interested researchers can tease out these further meanings such as 'prediction' or 'ability' or even more meanings through a qualitative analysis of specific examples. The reliability of

the framework increases when the number of sub-categories is smaller as they will be broader and will allow for a greater range of meanings.

The four sub-categories in Hyland's Text-Oriented bundles category, 'Transition', 'Resultative', 'Structuring', 'Framing', seem to be more transparent for categorization purposes than Biber et al.'s 'Topic Introduction' and 'Topic Focus' sub-categories which they group under Discourse Organizers. Indeed, Framing signals and Structuring Signals (e.g. *as shown in figure*) are classified as 'Referential Expressions' in Biber et al.'s (2004) system, but Hyland's placement in the TO category makes more sense because these bundles serve to organize the discourse for the reader and are thus compatible with the other two TO sub-categories. This 'text about text' seems to be in the right place in Hyland's (2008) system. Cortes (2004) also seems to think so since she categorized Framing bundles under Discourse Organisers.

The 'Special Conversation Function' category in Biber et al.'s (2004) framework is not mentioned by Hyland (2008); so presumably bundles belonging to this category were not found in his corpus.

Hyland's (2008) framework has some advantages over Biber et al.'s (2004): the sub-categorization seems to make more sense (e.g. Framing and Structuring Signals under Textual Oriented Functions), and the smaller number of categories makes Hyland's system easier to adopt for other studies aiming to do a functional analysis of LBs. However, Hyland does not provide many examples of LB functional analysis, nor does he acknowledge the multi-functionality of bundles as do Biber et al. (2006). A case in point is the LB *at the same time*, listed in the Location sub-category under RO bundles, which could just as easily be categorized as a 'Structuring Signal' under the TO function, as pointed out by Rezoug and Vincent (2018).

There are further challenges associated with applying functional frameworks. Different labels attributed by researchers to the same function or similar functions also add

to the confusion. For example, Ädel and Erman point out that the Focusing sub-function is labelled as Discourse Organizing by Chen and Baker (2010) and Cortes (2004), but Referential by Simpson-Vlach and Ellis (2010) and Biber et al. (2004).

Generally, there are problems with applying functional taxonomies to LBs because of the multi-functionality of bundles (Ädel and Erman 2012, Biber 2006: 139). Frequently, LBs fulfil different functions depending on the context and it is hard to decide on the function unless the context of use is examined. In corpus linguistic approaches, the volume of data can make this step prohibitive. This can affect the reliability of studies and lead to different results from the same data.

2.7.6. LBs and disciplinary variation

Many researchers (Biber 2006, Ebeling and Hasselgard 2015, Hyland 2008 and Scott and Tribble 2006) have agreed that there is a need to investigate disciplinary variation in the use of LBs because of their tendency to ‘occur and behave in dissimilar ways in different disciplinary environments‘ (Hyland 2008: 20). The tendency for specific multi-word sequences to be preferred by particular disciplines may have pedagogical implications for course material developers and as Hyland (2008) points out, the findings of these studies ‘may make depressing reading for commercial materials writers seeking to identify universals of academic writing‘ (2008: 11). Structurally complete LBs could be directly used by course-material developers but the findings from discourse analysis and corpus analysis can have other sorts of pedagogical implications; LB findings can be indicative of language preferences in a particular genre or kind of discourse which may be of indirect use in informing syllabus design. This section reviews studies that have focused on disciplinary variation in the use of LBs and the differences in LB use among groups of writers.

Hyland (2008) reports disciplinary variation in the range, function, structure, and proportion of LBs in Electrical Engineering, Microbiology, Business Studies, and Applied Linguistics, and variation between published writing and student writing (120 published papers, 80 doctoral theses and 80 postgraduate dissertations). Among the top 50 LBs in the four disciplines he investigates, over half of the items in each list were unique, indicating

that disciplinary variation is a major factor affecting distribution. From the functional perspective, RO bundles were more common in the Science and Engineering texts while the soft sciences preferred PO bundles. The highest frequencies of PO bundles were found in the Social Science texts, especially in research articles. The student texts rarely contained these type of bundles but when they did, they were always Engagement bundles. The findings of LB studies undertaken on expert genres such as RAs cannot be interpreted to mean that these features should be emulated by students since pedagogic genres have their own purpose and audience which most likely entail the occurrence of a different set of LBs from expert genres.

Rezoug and Vincent (2018) used Hyland's framework (2008) to analyse the structure and functions of LBs in the Algerian Corpus of Engineering (ACE), which consists of Masters dissertations in four sub-disciplines of Engineering (Power Engineering, Control Engineering, Computer Engineering, and Telecommunication Engineering) and comprises 594,599 words. However, they adopted stricter thresholds as criteria for including LBs: a minimum occurrence of 40 times per million words in at least 20% of the dissertations (as opposed to Hyland's 20 pmw and 10% dispersion rate).

Rezoug and Vincent report variation from Hyland's results in terms of preferred LB structures: using their rather strict thresholds, passive and anticipatory *it* pattern LBs were not as frequent as other types of LBs in ACE. Perhaps these features are associated with more proficient writing; Hyland (2008) also reports that they were least frequent in his least proficient sub-corpus (Masters level).

Disciplinary variation in terms of function is also reported by Rezoug and Vincent (2018). As in Hyland's corpus, RO bundles were the most frequent followed by the TO and then by the PO bundles, but higher frequencies of RO bundles related to quantification and description were found in Telecommunication Engineering than in the other engineering sub-disciplines.

This section indicates that disciplinary variation is often found in the use of LBs and also raises the question of genre variation in LB use. Little research, if any, has been undertaken to explore LB variation across genres in a specific discipline.

2.7.7. Variation in LB distribution in Expert and Novice texts

The extent to which LB analysis can reveal differences between expert and novice writers and between native and non-native writing has been the focus of numerous studies of LBs (Ädel and Erman 2012, Chen and Baker 2010, Cortes 2004, De Cock 2000, Nekrasova 2009). Most of these studies recommend that student writers and NNS should attempt to imitate the styles of expert writers and NS seemingly, based on the assumption that the writing situations in all contexts are the same. It is not generally advisable for student writers to copy all the techniques of expert writers, as I pointed out earlier, since pedagogic genres have different purposes. The comparison of NNS and NS writing is also not valid, especially in the light of the argument that the students in lingua franca countries only need to acquire 'adequate' writing skills to produce successful assignments. It is useful to review studies in this area, however, in order to evaluate the extent of research that has been undertaken and to gauge the areas that are under-researched, especially with reference to EFL learners in the 'expanding circle'.

Cortes (2004) compared the LBs used by expert writers in journals to student writers at three levels (undergraduate lower-division, undergraduate upper-division, graduate). Like Hyland (2008), who compared Masters students, doctoral candidates and experts, writers at different levels of proficiency were compared in this study. Cortes reports that student writers rarely used the LBs preferred by the experts and that when they did use them, they used them for different purposes. However, it would appear that Cortes, like other researchers (Hyland 2008, Chen and Baker 2010), assumes that students need to use the same type of LBs as those found in published articles, not taking into account the generic differences between expert and student texts.

Differences in LB use between novice and expert writers as well as native and non-native writing were found by Römer (2009) who examined one linguistic feature (*introductory it*) in LBs in four sub-corpora of expert and apprentice texts by both NS and NNS writers in Humanities subjects. The expert texts consisted of journal articles while the apprentice texts consisted of argumentative essays and coursework by undergraduate and

graduate students. The *introductory it* pattern was first identified as an interesting feature to study (LBs and frames were retrieved and analysed to come to this decision) and later frames containing the *it is* pattern were retrieved from the four selected sub-corpora. Römer reports that learners' use of the introductory *it* pattern deviated from expert writing in terms of structure (e.g. *it is (ADV) ADJ to-infinitive pattern*), functions, and choice of adjective. For example, the learners tended to use 'extreme' adjectives (e.g. *amazing, wonderful*) and also a much smaller range of adjectives than that used by the expert writers. Römer (2009: 156) describes 'extreme' adjectives as 'more emotional…adjectives that are more characteristic of speech than of academic writing'. These findings support the results of Cortes (2004) that there are differences between expert and student writing. Although this study cannot be classified strictly as an LB study except for the first stage in the methodology where LBs were retrieved from a reference corpus, it demonstrates a methodology which can be adopted when seeking to examine a specific linguistic feature after the first stage of LB retrieval. Römer (2009: 159) also recommends that we need to 'respond to the observed deviations from the expert norm in general ELT and in EAP classes', although there are no suggestions as to how learners may acquire these 'expert' uses of the *introductory it* pattern. Although this study, like Cortes (2004), equates expert and student writing requirements, there are some relevant pedagogic findings such as the use of extreme adjectives by students. The findings of Chen and Baker (2010) are also in line with these studies in that they found student writing (L1 English and L1 Chinese essays in BAWE) to be similar in their use of LBs but different from expert writing (published journal articles and 2000-word book excerpts in the FLOB corpus in mainly hard science disciplinary areas). However, it should be borne in mind that essays are not like articles in the hard sciences.

In their study of student assignments, Chen and Baker (2010) further refined their list of LBs by eliminating LBs containing content words (e.g. *financial, non-financial*), proper names and word sequences that were part of the assignment brief. Removing content LBs most probably led to the exclusion of frequent LBs which may have been very useful for students to learn. The method of eliminating 'content' words (e.g. *the transmitter and receiver, signal to noise ratio*) was also adopted by Pan, Reppen and Biber (2016). These

researchers point out that eliminating content words made the final LB list shorter and more generic. The bases on which these content words were chosen is unclear and lessens the rigour of their methods.

If LBs were retrieved for the benefit of EFL learners who may want specific support for a particular discipline, including common disciplinary words and ways of expression, this method of eliminating content words would not be ideal. Some of the supposedly technical words, considered as content words, identified by both Chen and Baker (2010) and Pan, Reppen and Biber (2016) do not appear very technical (e.g. *financial, non-financial*). Secondly, specific expressions may have considerable pedagogical value for some groups of learners. For example, consider the expression 'signal to noise ratio'; if learners are required to express ratios frequently in their disciplinary discourse, and if they do not possess the linguistic devices to express this concept, it would not be ideal to remove this particular multi-word sequence which can be exploited to raise learners' awareness of collocations around the word 'ratio'. Neither of these studies report consulting with the subject lecturers or EAP teachers before taking this step of removing 'content' words from their LB list.

Chen and Baker (2010) report that the experts used double the number of LBs containing NPs while only half as many discourse organizers as the students. This finding could be attributed to the inclusion of excerpts rather than complete expert texts in their corpus, as discourse organising LBs may have been excluded from their corpus because of its design. There is also the possibility that student genres require more discourse organizers due to their purpose (to demonstrate knowledge). Chen and Baker's study also reports differences in the LBs used by L1 Chinese students and L1 English students. For example, the Chinese students tended to use fewer hedging and epistemic markers in their LBs compared to the native speakers.

Chen and Baker (2010) claim another important finding: that it is proficiency rather than native language that is the most important factor that determines the use of LBs. They found that both NS and NNS students tended to avoid the phrasal style of writing that experts seem to prefer. These researchers, however, point out that their texts representing expert

writing were mostly from the hard sciences. According to Gardner, Nesi and Biber (2018), writing in the Humanities (even in the advanced levels) shows more clausal complexity and since the student texts in Chen and Baker's (2010) corpora were representative of both the Humanities and the Hard Sciences, it is difficult to conclude that proficient writing (in all instances) is more phrasal and that students should emulate this phrasal or 'expert' style. I once again point out the incomparability of pedagogic genres to expert genres such as book excerpts and RAs, since the differences between expert and student writing are too large, in terms of completeness, disciplinary area, and genre.

Future studies will need to be comparable in terms of discipline and genre to reach conclusive findings, though it would be nearly impossible to find experts attempting student genres. The solution could be to create corpora containing expert genres reproduced by students according to assignment briefs which mimic real-world genres such as Journal Articles or Magazine Articles. Chen and Baker (2010: 44) conclude that both the NS and NNS students would benefit from familiarity with expert users' LBs to 'achieve a more native-like style of academic writing'. Exactly how this should be done is not made clear. However, this study does serve to demonstrate the weaknesses of LB studies that are not based on comparable corpora.

2.7.8. LBs in Native Versus Non-native writing

In contrast to studies that have focused on differences in LB use between groups of writers with different levels of proficiency, others have compared LB usage in terms of native speaker status (i.e. NS vs. NNS). Ädel and Erman (2012) examined four-word LBs in their corpora of native English and Swedish L1 student essays in the discipline of Linguistics. They found that native speakers used a wider range of LBs, and used them more frequently (twice as often). Native speaker student LBs were also more likely to exhibit the linguistic features characteristic of more advanced writers, such as unattended *this,* existential *there,* and passive constructions, and items associated with hedging. Ädel and Erman also report that the non-natives made more informal lexical choices (e.g. *easy, hard*), and used a limited range of verbs in comparison with the native speakers. However, these findings regarding the greater frequency and range of LBs in native writing may not be conclusive because of

the differences in the nature of the essays produced by the two groups: the native writers' texts discussed published research while the non-native writing mainly involved analysis of empirical data.

Chen and Baker (2010) also report that the number of different LBs and the overall number of LBs are the greatest in expert writing and the lowest in NNS texts. However, the literature is rather inconclusive regarding this point because of the incomparability of the corpora they used, in terms of genre and discipline.

Pan, Reppen and Biber (2016) compared LB use between L1 English and L2 English professional writers and report significant findings about the differences between the two groups. This study is an extension of Chen and Baker (2010) and Ädel and Erman (2012) that compared LB use between L1 English and L2 students. The two corpora used by Pan et al. consisted of RAs in the field of Telecommunication written by L1 English and L1 Chinese professional writers. Pan, Reppen and Biber's study sought to remove some of the confounding variables that might have distorted their findings: both corpora were from the same discipline (Telecommunications) and contained the same genre; they were also of the same size, and the L2 texts were all authored by professional writers from the same L1 background (Chinese). Pan, Reppen and Biber (2016) report that the L2 writers in their study used fewer types and tokens in LBs as compared to their native counterparts.

Another finding of Pan, Reppen and Biber (2016) was that the non-native writers preferred LBs with verb phrases while the L1 English group favoured LBs with noun and prepositional phrases. This finding should be considered alongside others that have found that experts and more advanced writers prefer LBs with noun and preposition phrases (e.g. Chen and Baker (2010), especially in the hard sciences (Gardner, Nesi & Biber 2018 and Staples et al. 2016). Clearly, the pedagogical implications of these studies are that both novice academic writers and L2 students share some challenges in adapting to academic writing such as learning to use a phrasal style to condense information.

2.7.9. Association of LBs with Moves

Other insightful studies of LBs include those that have attempted to link LBs to specific moves and steps in a genre. Cortes (2013) examined LBs of various lengths (4 to 9) in a corpus of RA Introductions from various disciplines. Among her findings were the presence of long LBs (more than 6 words, e.g. *the remainder of the paper is organized as follows*) which tended to associate with one particular step in a move. However, since the corpus used in this study consisted of many different disciplines and no attempt was made to investigate LB variation among disciplines, it provides no insights into disciplinary variation in LB use.

Cortes (2013) points out that the association between these LBs and moves would be useful for students participating in genre-based writing lessons. This is an interesting methodological model that can be adopted by researchers who are concerned with describing student work for pedagogical purposes. As mentioned previously, it is risky to equate expert genres with student genres and conclude that students should try to imitate the way experts write if their communicative purposes are different. Cortes (2013) offers a partial solution; instead of trying to accomplish the impossible task of finding expert texts in typical student genres, corpora containing proficient student assignments (e.g. BAWE) could be divided into sub-corpora in terms of their rhetorical stages. There is pedagogical value in linking moves and steps to LBs as students can be made aware of the phraseology employed to achieve different communicative purposes.

2.7.10. Pedagogical Implications

The studies discussed above have established that speakers and writers often repeat the same fixed strings of words, and that there is a tendency for specific LBs to be preferred in particular disciplines and genres. Although most of the studies comparing L1 and L2 and expert and non-expert writing were methodologically flawed in some way, as pointed out above, there are indications that both L1 and L2 student writers face challenges in using word combinations appropriately. The ensuing section will now summarize and weigh the pedagogical implications relating to LBs arising from these studies.

The first concern regarding the recommendations of those studies which advocate direct 'teaching' of LBs is their questionable 'teachability'. It is generally difficult to teach and learn LBs, as most of them are incomplete phrases and are not perceptually salient. According to Biber et al. (1999), LBs are not meant to be directly transferred to the classroom, rather they add one more dimension to the description of language in a register. However, Hyland (2012: 165) argues for the 'teachability' of LBs by claiming that 'It is possible … for bundles to be taught in EAP classrooms, although to date very little by way of practical applications has been published'. He suggests that the LB is an independent 'linguistic construct' which is 'teachable'.

An attempt at using LBs to inform teaching has been made by Ellis, Simpson-Vlach and Maynard (2010) who compiled an 'Academic Formulas List' containing spoken and written LBs and phrases (e.g. *weight of the, it is obvious that, I have a question*) drawn from various academic sub-corpora (BNC (the British National Corpus), MICASE (the Michigan Corpus of Academic Spoken English), and Hyland's corpus of research articles). The extent of the usefulness of such a list of academic multi-word sequences to support student writing needs to be examined, since the corpora they used consisted of spoken English and published academic prose, which are different from the genres of student writing. However, in order to make their list of academic formulas more meaningful, they used not only statistical measures but also Mutual Information scores and consulted experienced EAP instructors and language testing experts.

Another issue that needs to be discussed in this context is that LBs vary according to discipline as some studies (e.g. Hyland 2008 and Rezoug and Vincent 2018) have shown. Genre variation is also a factor that needs to be considered while selecting LBs that may be useful for learners to be familiar with. There is thus a need for studies of LBs drawn from genres belonging to the same discipline which will narrow down the range of LBs required by students, especially those in EFL contexts.

The methods adopted to raise learners' consciousness about LBs cannot be restricted to conventional methods such as lists to be memorized, but need to follow a more 'in context'

approach where the LBs are not isolated from the surrounding text. Rezoug and Vincent's suggestion (2018) to use LBs as a starting point and cluster them together by function can also help point to recurrent phrases / phraseologies which might then be taught. This is an improvement on the 'Academic Formulas List' (AFL) approach because the lists would include phrases which are perceptually salient and which also have distinct discourse functions (AFL contains multi-word strings that are not perceptually salient such as *shown in figure, as a result of, to determine whether*).

The LB studies reviewed in the preceding sections can help with the descriptions of student genres and research articles written by experts. Thus these findings may have pedagogical value, but this will be indirect. They reinforce the need highlighted by Hyland (2012) to make writing classes discipline-specific. They may also have pedagogical value for course book developers and EAP instructors who can think about practical ways to incorporate these findings into their lessons through using concordance lines or other activities to raise awareness of the limited number of LBs that are complete phrases.

The similar findings of Hyland (2008) and Rezoug and Vincent (2018) regarding the absence of passive structures and anticipatory *it* patterns in student genres, might, after further investigation, have pedagogical value since they can alert EAP course book designers and instructors to learners' lack of familiarity with these structures. However, it may be that the absence of passives and anticipatory *it* in the Masters dissertations in Engineering analysed in these studies does not tell us anything about learner deficit, but rather tells us about the register of Engineering dissertations, which are generically different from research articles. This can be further explored on the basis of student corpora representing a substantial number of texts from a single level of study and one engineering discipline, so that genre, discipline and level of study effects can be isolated.

2.7.11. Implications of LB Studies

Although there are methodological flaws and some degree of conflict in the results reported by these studies, they have all pointed out systematic differences in the use of LBs across groups of writers and disciplines. Two main types of analysis have been conducted in these

studies: structural categorization and functional categorization. The structural analyses of LBs suggest that structurally different LBs are preferred by students and expert writers; expert writers tend to use more phrasal bundles (containing noun and preposition phrases) while student and L2 writing tends to favour verb phrases (Chen & Baker 2010, Pan, Reppen & Biber 2014). These findings may be correlated with those of recent large scale studies using corpus linguistics methods that have found that students at higher levels of study tend towards more phrasal complexity in their writing as compared to less proficient writers (Gardner, Nesi & Biber 2018, Staples et al. 2016). Register, genre, and disciplinary variations have been established in the distribution of structural LBs. Biber et al. (1999) have demonstrated the usefulness of analysing LBs by their structural types and identifying associations of structural types of bundle with particular registers.

The second type of analysis, involving the classification of LBs on the basis of their discourse functions, has also yielded some consistent findings traceable across most of these studies. The findings of these studies (Chen & Baker 2010, Hyland 2008 and Rezoug & Vincent 2018) can be extended by further research that examines and compares LBs found in genres within a single discipline.

As I have pointed out earlier, there may be several methodological problems affecting the validity of the findings: lack of comparability of the corpora used for the analysis in terms of genre and size, the different structural and functional taxonomies used to analyse LBs (Biber et al. 2004, Hyland 2008), the varying frequency cut-offs and dispersion rates, and the different LB identification procedures adopted by linguists.

The functional categorization of LBs using Biber et al.'s (2004) framework has been shown to be problematic; apparently the only way to identify their function would be to refer to the expanded contexts of these LBs. As there is a strong possibility that there will be differences of categorization across different researchers/analysts even if the surrounding text is examined, such analyses will continue to be 'problematic' and not reliable (Ädel and Erman 2012).

Findings from studies comparing LB use by L1 and L2 writers (Ädel & Erman 2012, Chen & Baker 2010) reveal some of the general features of the language used by L2 writers: they can alert researchers to areas in need of improvement in L2 writing such as, for example, the use of noun and preposition phrases to achieve information-dense writing, and the use of cautious language. The findings from these studies are only indicative, however, due to the methodological defects pointed out earlier such as the fact that the corpora being compared are often different in terms of genre and discipline.

The studies discussed above have established the pervasiveness of recurrent strings of words in writing, and have demonstrated the usefulness of representative and specialized corpora in the description of registers and genres. The range of genres explored are, however, limited. For instance, the types of LBs in specific genres written by undergraduate students in particular disciplines is an unexplored area. It would be useful for more such studies to be conducted considering that millions of students worldwide are entering English-medium tertiary education, a large proportion of whom are non-native speakers struggling to meet the English academic writing demands of their disciplines. The insights gained from these types of studies can inform ESP course materials and delivery.

To treat LBs as the 'building blocks' of discourse (Biber and Barbieri 2007, Biber 2006, Conrad and Biber 2004, Hyland 2008, 2012 and Chen and Baker 2010), might be overstating their importance. Nevertheless it seems reasonable to claim that studying recurring multi-word sequences in a corpus of texts representing a genre or discipline can help us with the linguistic description of these texts.

2.8. Phrasal Complexity in academic writing

Grammatical complexity is an important characteristic of proficient academic writing. Studies have demonstrated that this quality of successful university writing is often achieved through the use of complex phrasal structures (Gardner, Nesi and Biber 2018, Staples et al. 2016). Phrasal complexity is realised through the use of dependent phrases and noun modification. This review will discuss complex nominals and nominalizations (components of nominal structures), because although findings from recent corpus studies confirm their

prominent role in academic writing only a small number of studies that have focused on these features in student writing. The sections below will discuss these features and review the most relevant studies in this area.

2.8.1. Complex Nominals and Nominalization

A nominal group is considered to be a single grammatical unit with the Head noun being its most important component. The addition of multiple pre and post modifiers to the Head (to form what I will refer to here as 'noun strings') increases its complexity. Researchers include nouns, attributive adjectives, nouns as nominal pre-modifiers, *of* genitives, and other prepositional phrases and embedded clauses as part of modifying elements constituting complex nominals (Bloor and Bloor 2013, Staples et al. 2016). Complex nominals, or multi-word noun phrases, are formed as a result of moving post nominal structures such as prepositional phrases or relative clauses which are located to the right of the Head noun, towards the left to pre-modifier position. This causes the prepositional marker and words that signal semantic relationships among the components of a sentence to be removed. Montero (1996) attributes this preference for nominal structures in scientific writing to three main reasons: linguistic economy, a desire for new forms of expression, and the need to express new scientific concepts.

Linguists categorise nouns as nominalizations when they are derived from adjectives or verbs. The derived forms usually have suffixes such as *-tion, -ity, -ment* and *-ness* (Biber et al. 1999), but verbs can also be converted with a 'zero suffix' and no alteration in their form (e.g. *comment, mention*). Nominalization performs the function of 'reification' (Biber et al. 1999, Halliday and Matthiessen 1999, 2004), and occurs when processes (congruently expressed through verbs) and properties (congruently expressed through adjectives) are metaphorically re-construed (Halliday and Matthiessen 2004), for example when the process *adjust* becomes *adjustment* and the quality *safe* becomes *safety*. This type of 'non-congruent' or metaphorical use of language facilitates the expression of extended meaning, and 'is the single most powerful resource for creating grammatical metaphor' according to Halliday and Matthiessen (2004: 656). Halliday and Martin (1993, 1999, 2004) regard it as one of the

primary ways to achieve the lexically dense and nominal style of academic discourse, and Halliday and Matthiessen (2004: 657) speak about the 'prestige and power' bestowed on writers through the reification of their concepts.

According to Halliday (2004: 657) the use of nominalization 'probably evolved first in scientific and technical registers', to construct arguments and 'hierarchies of technical words'. Nominalizations can sum up 'a fairly complex argument' (Halliday 1993: 60) which is already 'given'. The ability of nominalizations to express technical terms, build knowledge, and lend cohesion to text is highlighted in Halliday's description.

2.8.2. Studies of Nominal Compounds and Nominalizations

A number of recent large-scale studies of university student writing (Biber, Gray and Poonpon 2011, Gardner, Nesi and Biber 2018, Staples et al. 2016) have reported on the prominent role of nominal groups and nominalizations (see Table 2.7). Their findings have relevance in the current discussion because they provide empirical evidence of the type of complexity that is favoured in proficient student writing.

Individual studies that have investigated these phrasal features are reviewed in the next section. Table 2.7 summarizes the main features of these studies.

Study	Corpus Size/Number of texts	Additional details of corpora
Charles (2003)	• Politics: 200 thousand words (8 texts) • Material Sciences: 300 thousand words (8 texts)	• MPhil Theses • Doctoral Theses
Aktaş and Cortes (2008)	• Graduate writing: 66, 459 words (28 texts) • Published texts: 721,553 words (166 texts)	• Art and Design, Biology, Computer Science, Economics, Environmental Engineering, Physics and Astronomy • The same 6 disciplines represented in both sub-corpora
Biber, Gray and Poonpon (2011)	• Academic writing corpus: 2,939,000 (429 texts) • Face to face conversation: 4,175,000 (723 texts)	• Academic writing (research articles from biology, education, history, medicine, psychology) • Face to face conversation
Biber and Gray (2013)	• Academic prose: 5.2 million • Newspaper reportage: 250,000 • Fiction: 12 million • Drama: 80,000	• academic prose (science research, non-science research, popular science) • newspaper reportage • fiction • drama
Parkinson and Musgrave (2014)	• Corpus 1: 13, 711 words (21 texts from the EAP group) • Corpus 2: 12, 577 words (16 texts from the MA group)	• Corpus 1: argumentative essays • Corpus 2: disciplinary assignments with literature survey component
Flowerdew and Forest (2015)	• Natural Sciences: 288, 242 words (52 texts) • Social Sciences: 325, 272 words (49 texts)	• Lectures, journals, textbooks • Natural Sciences: Biological Science, Chemistry, Engineering Science, Physics and Ecology • Social Sciences: Economics, Politics and International Relations, Law, Sociology and Business Studies
Staples et al. (2016)	BAWE sub-corpus: 4,480,371 words (1948 texts)	Student assignments from the disciplinary groupings of Arts and Humanities, Social Sciences , Life Sciences and Physical Sciences

Table 2. 7. Summary of studies of phrasal features

Biber, Gray and Poonpon (2011) emphasize the prominence of phrasal structures in advanced academic writing and propose a developmental progression index of increasing noun phrase complexity. This study offers empirical evidence of the complexity measures that may be useful in evaluating academic writing. The traditional method of measuring complexity in academic writing by counting the length of the clause and the number of subordinate clauses is critiqued, and a 'principled basis for the selection of complexity measures' (32) is offered instead. Biber, Gray and Poonpon compared two corpora belonging to the written and conversational registers to investigate whether extensive subordination is a feature of advanced writing, and uncover the linguistic devices used to achieve complexity in written academic discourse. The frequency of occurrence of three main grammatical structures were compared (finite dependent clauses, non-finite dependent clauses, and non-clausal dependent phrases). Based on their findings, Biber, Gray and Poonpon propose a progression route from finite dependent clauses towards non-finite dependent clauses, and finally towards the dense use of non-clausal phrasal structures in advanced academic writing. These phrasal features include prepositional features, nouns as nominal pre-modifiers, and attributive adjectives.

Although Biber, Gray and Poonpon only examined RAs and conversation in their corpus, they propose that embedded phrasal structures are acquired quite late in a student's academic career because such kind of complex writing is required only when a student is initiated into a disciplinary discourse community. The study has added to our knowledge about the features of published academic writing, but more studies of the phrasal features preferred in different disciplines and genres are required for a deeper understanding of the role of the situational variables which may affect students' choice of linguistic devices.

The validity of the developmental progression index proposed by Biber, Gray and Poonpon was tested by Parkinson and Musgrave (2014), who analysed 3970 noun phrases in argumentative texts written by students enrolled on an EAP course preparing for graduate study, and 3390 noun phrases in disciplinary assignments with a literature review component written by graduate students of Applied Linguistics. They found significant differences between the two groups in terms of certain grammatical features used to modify nouns, thus supporting Biber, Gray and Poonpon's idea of a developmental progression index. For

example, they found that attributive adjectives, which are placed earlier in the index, were more frequently used by the EAP group, while pre-modifying nouns and prepositional phrases with phrases other than *of* (placed later in the progression route proposed by Biber, Gray and Poonpon) were used much more frequently by the MA group.

Parkinson and Musgrave (2014) is insightful because of the detailed qualitative analysis of the texts. The findings lend strength to the argument that certain nominal group components are acquired by students later in their academic journey. However, this study used only a small number of texts (see Table 1 for details) and compared only two different kinds of writing (argumentative essays set in an EAP class and disciplinary assignments in Applied Linguistics), ignoring the possibility that genre variation may account for the different noun modifying features used, as can be seen from Biber and Gray's study reviewed below.

Biber and Gray (2013) report a historic increase in the use of nominal groups, and a corresponding decrease in verbs and clausal features in academic prose. They created two corpora, of academic and non-academic texts from the years 1900, 1925, 1965, 1985 and 2005. They divided their academic corpus into three: technical science writing, non-technical research writing in disciplines such as Education, Psychology and History, and popular science writing. The non-academic texts came from newspapers, drama and fiction. Biber and Gray's most notable finding regarding technical science writing was the increase in nouns, and nouns as nominal pre-modifiers, accompanied by a general increase in other nominal modifiers such as attributive adjectives and preposition phrases. There was a dramatic increase in the use of nouns as pre-modifiers in the two information-dense registers of academic prose and newspaper writing from the end of the twentieth century onwards, the shift being the most prominent in science research articles. The incidence of nominalizations was much higher in the non-technical sub-corpus than in the science sub-corpus, and in the popular science articles compared to specialist science research texts. This implies that the distribution of these features is dependent on the genre and therefore these findings need to be verified by further research into specific genres in particular disciplines.

2.8.3. Recent large scale corpus studies on phrasal complexity features

Staples et al. (2016) examined the development of phrasal complexity in L1 English writers across four levels of study (undergraduate to taught Masters level), and investigated the influence of disciplines and genres on this process. A sub-corpus of the BAWE corpus comprising assignments written by L1 English students across the disciplinary groups of Arts and Humanities, Social Sciences, Life Sciences and Physical Sciences was used in their study. Phrasal features used to modify or elaborate nominal groups were found to be more frequent in the hard sciences. Pre-modifying nouns were found to be the most frequently used phrasal feature at the highest level of study with an increase of about 40% from Level 1. Among all the phrasal features, pre-modifying nouns also showed the highest interaction effect of level and discipline; Explanations and Case Studies have more of these nouns than any other genre. In BAWE, Explanations are frequent in Life and Physical Sciences, while Case Studies are frequent in Business and Health disciplines.

Nominalization use also increased between Levels 1 and 2 and then again between 3 and 4, accounting for an increase of 18% across levels. They occurred most frequently in Essays in the Social Sciences. However, since all the disciplines were grouped into four broad disciplinary areas, the behaviour of nominalizations in specific disciplines was not investigated.

The most important finding of this study is the increasing phrasal complexity of student writing across levels, which is accompanied by declining clausal complexity. The identification of the pre-modifying noun as the most significant phrasal feature of proficient university writing deserves further investigation, as situational variables such as discipline and genre need to be considered. The findings of Staples et al. (2016) regarding the prevalence of nominalizations in the Social Sciences support findings from previous studies (Biber, Gray and Poonpon 2011, Biber and Gray 2013). However, their findings diverge from the claims of other researchers (Banks 2008, Halliday 2004) who regarded this feature as a prominent linguistic device in scientific and technical writing (see 2.8.1).

In another large-scale study of university student writing, Gardner, Nesi and Biber (2018) provide a linguistic description of the texts in the BAWE corpus with reference to the situational variables of genre, level of study, discipline and disciplinary group. They found that Physical and Life Sciences texts across the levels, and genres such as Methodology Recounts, Design Specifications and Case Studies, are characterized by relatively high frequencies of pre-modifying nouns, passives, action verbs, and common, quantity and concrete nouns. Texts in the Social Sciences were often characterized by the use of long words, nominalizations, attributive adjectives, and abstract nouns, while these features were not prevalent in the Physical Sciences. These texts also tended to be more clausal than the assignments in the Physical and Life Sciences. Gardner, Nesi and Biber (2018) point out that there are two types of density: the compressed procedural information typical of Science reports (pre-modifying nouns, common nouns, passives, action verbs, concrete and quantity nouns) and the type of density characterized by linguistic features such as nominalizations and attributive adjectives, found in the Social Science texts. This ties together the findings of all the previous studies (Biber and Gray 2013, Staples et al. 2016) regarding the two different types of density in academic writing.

However, one should consider that different types of assignments can be set within the same discipline. For example, within Engineering some assignments might be 'softer' than others. BAWE Case Studies in Engineering may tend be more similar to writing in the Social Sciences, as the module titles reveal (e.g. *Economics and the Structure of Industry, Engineering Business Management 1*).

Gardner, Nesi and Biber's study maps not only disciplinary groupings, but also genres and specific disciplines to clusters of linguistic features. Other studies can build on these mappings by examining them in specific disciplines and genres. For example, researchers interested in exploring the linguistic devices used in Methodology Recounts in Engineering can identify them with reference to the findings reported in this study, and with reference to the situational variables mentioned in the previous paragraph. This study can thus, as the authors claim, be used to inform the teaching of a 'common academic core' as well as to facilitate the design of discipline-specific courses.

2.8.4. Signalling Nouns

The linguistic device called ‘Shell Nouns’ or ‘Signalling Nouns’ (SNs) deserves separate treatment within a comprehensive overview of the different functions of nominalizations.

Schmidt (2000: 4) defines SNs as ‘an open-ended functionally-defined class of abstract nouns that have to varying degrees the potential for being used as conceptual shells for complex, proposition-like pieces of information’. They are defined by Flowerdew and Forest (2015) as ‘abstract nouns which are non-specific in their meaning when considered in isolation and which are made specific in their meaning by reference to their linguistic context’. Ivanic (1991: 109) points out that an SN has ‘both a constant and a variable meaning, and the variable meaning is dependent on the context in which it is used’. These definitions highlight the ability of these nouns to gain meaning by encapsulating complex information contained in the surrounding text.

The additional meaning taken on by these cohesive devices from the surrounding discourse is discussed by Schmidt (2000: 13-14), who explains that deciding whether a noun is an SN ‘does not depend on inalienable characteristics inherent in the noun, but on its use’. He characterises the three functions of this type of noun as follows: ‘the semantic function of characterizing complex chunks of information, the cognitive function of temporary concept formation, and the textual function of linking these nominal concepts with clauses which contain the actual details of information’. SNs thus carry indexical meaning whilst also functioning as cohesive devices; the semantic relationship of SNs to text is realized through a process of ‘agnation’, whereby the meaning of the SN is equated with a clausal component or a nominalization preceding or following it. Sinclair (2004) used the term ‘encapsulation’ with reference to the anaphoric use of signalling nouns, and the term ‘prospection’ with reference to their cataphoric use.

Although some frequently used SNs such as *problem*, *fact* and *idea* are not nominalizations, Flowerdew and Forest (2015) point out that ‘nominalisations …are productive processes for the creation of new SNs’. They can encapsulate mental processes

(*assumption*, *prediction*), verbal activities (*discussion*, *argument*), modalised facts (*probability*, *possibility*) and circumstantial facts (*condition*).

SNs have been given various names in the literature: *signalling nouns* (Flowerdew 2003, Flowerdew and Forest 2015), *Type 3 vocabulary* (Winter 1977), *advance and retrospective labels* (Francis 1994), *carrier nouns* (Ivanic 1991), *metalanguage nouns* (Winter 1992), and *shell nouns* (Schmid 2000). Francis further categorizes head nouns acting as retrospective labels according to their derivations from speech acts (*accusation, announcement*); verbal activities (*implication, eulogy*); metalinguistic 'text' nouns (*argument, point, distinction, expression and quotation*); and cognition nouns (*analysis, viewpoint*).

The prevalence of SNs in academic writing makes them the subject of increasing scholarly interest, as evidenced by their recent book-length treatment by Flowerdew and Forest (2015) which claims to focus on their discourse rather than sentence-level realization. Flowerdew (2003) described the cohesive functions of signalling nouns by exemplifying their anaphoric and cataphoric use in two corpora of approximately a million words each, one consisting of biology lectures and the other of texts taken from the prescribed biology textbook. Signalling nouns were found to occur more frequently in the textbooks than in the lectures, with both genres displaying distinct preferences for certain words. Flowerdew and Forest (2015) based their study on two relatively small sub-corpora in Natural Science (NTS) and Social Science (SS) comprising about 600,000 words in total, marked up syntactically and semantically. Each sub-corpus contained five disciplines, each with a roughly equal representation of three genres: RAs, textbooks and lectures.

Flowerdew and Forest (2015) found that SNs were frequent in their corpus with an average occurrence of one in 37 words. They also found that:

- SNs are more frequent in SS compared to NTS texts (accounting for 70% of occurrences).
- The RA genre has the highest frequency of SNs.

- *Case*, *way* and *problem* are the most frequent SNs in the corpus (these are not nominalizations).
- Anaphoric SNs are more frequent than cataphoric SNs.
- Certain patterns are favoured by certain genres (e.g. textbooks: SN + appositive) and disciplinary groupings (e.g. SN + be + that clause).

The inherent fuzziness of the SN concept is demonstrated by the researchers' description of the challenges they faced in identifying the lexical realizations of SNs. It was difficult to map technical words and bivalent clauses (when there are two clauses following the SN and only one is equative to the SN) whose meanings are compacted in the SNs. The limitations of their study, as Flowerdew and Forest themselves acknowledge, include the questionable accuracy of their annotation scheme, the small size of their corpus, the low likelihood of replicability, and the failure to qualitatively analyse the examples provided. Nevertheless, these findings can be extended by studies that explore the influence of genre and discipline specificity on the distribution of Shell Nouns, especially in pedagogic genres so that they can inform the teaching and learning of disciplinary writing.

Charles (2003) used corpus techniques to study the use of shell nouns in two corpora of theses in Politics and Material Sciences. The Politics texts contained twice as many stance nouns, a finding which she attributes to the nature of the activities of this discipline which are mainly interpretations based on other texts in written form. She draws attention to the use of 'retrospective labelling' (anaphoric reference) to achieve cohesion in the process of building arguments in both the corpora. This study, however, did not attempt to compare patterns of SN use as many other studies do.

Aktas and Cortes (2008) examined SN use in a corpus of 28 NNS graduate papers and a corpus of 166 published RAs. They identified seven lexico-grammatical patterns associated with signalling nouns (see Table 2.8).

Table 2. 8. SN patterns found in student and published RAs (Aktaş and Cortes 2008)

Aktaş and Cortes identified the patterns favoured by each group and the functions for which they are used. Amongst the NNS texts they reported more frequent use of the 'Determiner + Noun' pattern used anaphorically for a linking function, and in the published texts they reported a predominance of the cataphoric 'Noun + Clause' structure used for the function of characterization. They also found that professional writers used SNs for temporary concept formation more frequently than graduate students. Differences in writing style are expected when graduate students are compared with academics, due to their different level of disciplinary knowledge, their purpose and readerships.

Nesi and Moreton (2012) explored the differences in shell noun use between NS and NNS students in the BAWE corpus. They retrieved and examined those shell nouns in BAWE which were also identified by Hinkel (2004) and Flowerdew (2003). The most frequently used set of nine nouns in the BAWE corpus were examined for patterns identified by Aktaş and Cortes (2008) and it was found that 'N + cl' (Noun + post-nominal complement

clause), 'th- + N' (Demonstrative determiner + Noun) and 'a/the + N' (Article + Noun) were the most frequently occurring patterns. They also found that in BAWE both NS and NNS students preferred the 'N + cl' pattern and that both used SNs redundantly. The pedagogical value of such studies is that EAP teachers and students can be made aware of the need to use SNs appropriately.

To sum up, recent studies have established the situational variables of discipline and genre as predominant factors influencing the choice of linguistic devices that contribute to the complexity of proficient student texts. Much more is now known about the specific phrasal features that contribute to this complexity in a disciplinary group, genre or discipline. Nominalization and noun pre-modification are two phrasal devices that feature quite often in the literature on student writing. This line of research needs to be continued for the sake of millions of students worldwide, especially L2 writers who require research-informed pedagogies to help them conform to the conventions of their discourse communities.

2.8.6. Complex Nominals and Nominalizations: Complexity or Ambiguity?

Having discussed the predominance of noun compounds and nominalizations, and some of their functions in academic writing, some space should be given to discussions of how they can make texts too complex. Linguists such as Montero (1996), Ruiz (2006) and Varantola (1984) have pointed out that the use of continuous noun strings in scientific writing can cause ambiguity since the semantic relationships among the components of the noun phrase are sometimes left to the reader to process. These linguists recommend that the addition of more than two pre-modifying elements to the head noun is to be avoided. Montero, however, found that more than two elements in noun phrases was relatively uncommon in his data (4325 noun phrases taken from Computer Science sources). Most university websites on writing instruction (University of Wisconsin-Maddison, Purdue Online Writing lab) warn students against stringing nouns together because of these pitfalls.

It is interesting to note that nominalization use in the Social Sciences has also come under attack. Like noun phrases, nominalizations increase the complexity of texts, and can make them inaccessible to disciplinary outsiders (Bloor and Bloor 2013, Montero 1996).

Billig (2013) has been particularly scathing in his attack on established Social Science writers who he accuses of deliberately using nominalizations to confuse the reader. He contends that by reifying the theories that they propound (*massification, globalization*) scientists aim to make them appear more profound. He also attributes an ulterior motive to this linguistic predisposition: to hide a lack of clarity in thinking and to keep out those who are not familiar with the jargon of the discipline. He suggests that in this era of super-specialization, communication between even neighbouring disciplines breaks down because of this habit. The coining of new terms using nominalizations even when existing terms can be employed or modified to express new concepts is another accusation levelled at social scientists; the unprecedented rate of academic publications due to the pressure to publish may cause, according to Billig, a lack of ideas and a resort to 'big ifications and izications' (2013: 115). Billig believes that nominalization is not necessary for social scientists, but that its adoption by science writers is forgivable since they do not have alternatives. Billig's criticism of the 'nouny writing' of social scientists becomes especially relevant in the light of the findings of Gardner, Nesi and Biber (2018) regarding the predominance of nominalizations in students' Social Science texts.

Critical Discourse analysts such as Wodak and Meyer (2010) and Bloor and Bloor (2013) have also discussed how nominalizations can be manipulated to distort meanings and to hide the agents of actions, although Bloor and Bloor's (2013) claim that it is primarily grammatical metaphors that lend lexical density to academic texts ignores other contributing features such as noun pre-modification..

2.9. Chapter Summary

A discussion of the techniques of genre analysis (Swales 1990, 2004), of Lexical Bundle analysis (Biber et al. 1999, Hyland 2008) and of phrasal complexity features (e.g. Gardner, Nesi and Biber 2018, Staples et al. 2016) have framed this chapter. This chapter has attempted to evaluate research on academic writing, with a specific focus on student genres, disciplinary variation, engineering writing, academic writing contexts in the 'expanding circle', and the phraseology and phrasal complexity features of written academic genres. I have reviewed genre identification studies that have used various methods such as surveys,

assignment prompts, and corpora of student assignments, information obtained from specialist informants and sometimes through a combination of all these methods. Disciplinary variation in academic writing has been explored by evaluating the studies that have used genre analysis and corpus techniques to reveal the similarities and differences in move structure and linguistic features across disciplines, genres and other situational variables. Phraseology has been discussed by examining the structural and functional frameworks developed and applied to study Lexical Bundles. The chapter concludes with an evaluation of the research on the phrasal complexity features in academic writing with a focus on noun strings, nominalisations and attributive adjectives.

Students and institutions positioned in 'expanding circle lingua franca' countries have unique needs with regard to academic writing support. I have argued for the need to prioritize clear and accurate writing rather than elegance in communication. These priorities should drive the research and pedagogy designed for supporting the needs of these EFL learners. This emphasis implies the acceptance of a restricted communicative code which is adequate or 'fit for purpose' for the communicative needs of these students.

One effective way of supporting the writing needs of these students would be to conduct more research based on corpora of texts collected from successful student writers in these contexts. The comparatively limited set of syntactic and lexical resources required to write successfully can then be identified. This approach will perhaps be more helpful than exposing students to an overwhelming array of linguistic features which many students will find impossible to learn to use.

To further narrow down the linguistic resources required by these learners and considering the findings of research on disciplinary variation (see Section 2.4), the way forward seems to be to create student corpora in specific disciplines so that students are familiarized with the generic structures and lexico-grammatical features of the genres within their own discourse community.

I now attempt to situate my own study of Omani student writing in Civil Engineering in the light of my discussions in this chapter regarding academic writing and the unique needs

of students in the 'expanding circle lingua franca' countries. I propose the creation of a single-discipline corpus of student writing consisting of texts collected from a single institution, to facilitate an in-depth linguistic characterization of the genres assigned within this single discipline.

I have chosen the Civil Engineering (CE) discipline because of the importance of clear and accurate writing in this field and also because I was able to collect a substantial and representative number of texts across all the modules and semesters of study selected. The texts will first be categorized into genres to identify the primary genres in CE. I will then use corpus linguistics methods to characterise these primary genres and to reveal generic variation among the genres assigned to CE students. The single-discipline design of my corpus will enable the isolation of genre characteristics. I am not aware of other studies located in the 'expanding circle' that have attempted to create such 'lingua franca' corpora in a single discipline in order to facilitate in-depth genre descriptions informed by genre analysis and corpus linguistics techniques.

The RQs posed for the purposes of this study are as follows:

1. What is the move structure of the two main genres in Civil Engineering?

2. What are the similarities and differences in the phraseology of the two genres?

2a. What are the structures and functions of the Lexical Bundles retrieved from the Case Study (CS) and Methodology Recount (MR) sub-corpora?

2b. How are the Lexical Bundles retrieved from the CS and MR sub-corpora similar or different?

3. What are the similarities and differences in terms of phrasal complexity and other nominal features?

3a. What are the similarities and differences in the frequency and function of nominalisations, attributive adjectives, and noun strings in CS and MR assignments?

3b. What are the noun types used in *the N of (the) N* pattern in CS and MR assignments?

Chapter 3

Methodology

This chapter describes the methodology adopted to attempt a linguistic characterization of the two main genres in undergraduate Civil Engineering writing at the institution where the research was conducted. A Corpus Linguistics approach has been used to examine the similarities and differences in phraseology and phrasal complexity features of the student assignments in these two genres. In addition, this study draws upon Swales' work on genre analysis (Swales 1990, 2004) which facilitates the description of the rhetorical sections of texts and the 'schematic structure' within these sections, consisting of moves and steps. Since text should not be isolated from context, an ethnographic dimension complements the findings obtained from the use of corpus analysis techniques.

The first section describes how undergraduate student assignments in Civil Engineering were collected and compiled into a corpus. It then discusses how the genres represented in the corpus were identified on the basis of the actual student texts, course documentation and on-going interactions with specialist informants. The detailed move structure of the two main genres identified in the corpus is also included in this chapter. The chapter then sets out to describe the corpus techniques used to retrieve the LBs, phrasal complexity features and the *N of (the) N* patterns from the two main genres identified in the corpus. The frameworks used to analyse the LBs and the semantic noun categories in *the N of (the) N* pattern are also discussed.

3.1. Civil Engineering writing

The student assignments were collected from the Civil Engineering (CE) Department at the research site in the Sultanate of Oman. I chose CE student assignments for three reasons. Firstly, a preliminary reading of the literature had revealed that corpora of student writing in engineering are limited to one or two relatively small sections of the MICUSP and BAWE corpora. The former contains only 31 texts under 'Civil & Environmental Engineering' while BAWE includes 238 assignments from 'Engineering' out of which only 9 assignments are

from Civil Engineering courses. Secondly, no corpus of Arab student writing in engineering exists to the best of my knowledge, except for the ACE corpus. Hence, in order to examine the assignments in this Arabic EFL context, the need to create a corpus of engineering student writing is clearly established.

Most importantly, I found that I had collected a substantial number of assignments in CE, submitted in a digital format, which seemed to be from a range of genres. Initial data collection from other disciplines which I had thought of including in this corpus revealed that many of the assignments were handwritten and would require transcription. By deciding to focus on one discipline, I have allowed myself space for an in-depth exploration of the unique characteristics of CE genres, whilst also satisfying the condition for case study research which 'seeks depth rather than breadth in its scope and analysis' (Mackey and Gass 2012: 96). By including other disciplines I would have broadened the scope of the study at the cost of 'depth'.

3.1.1. Context

This study draws on ethnographic approaches as the inseparability of text and context in academic writing research is well-accepted. Hammersley (1992, 2006) characterizes ethnography as an approach which draws on naturally occurring contexts (opposed to researcher-created conditions) to gain insights from an 'emic' or insider's perspective (rather than an 'etic' or 'outsider' researcher perspective). Hammersley and Atkinson (2007) emphasize that this approach requires sustained involvement of the researcher with the participants at the research site. The two approaches to ethnography emphasized by Lillis (2008: 362) in relation to studies of academic writing practices are 'lengthy or sustained engagement in participants' academic writing worlds, and the collection and analysis of a range of types of data in order to build holistic understandings'. She points out that 'talk around texts' would help the researcher gain insights into contextual influences on the text which are 'beyond the text' (361).

This study fulfils the conditions of an ethnographic approach as interpreted by Lillis (2008) and Hammersley and Atkinson (2007): my prolonged engagement of nearly 15 years with the students and lecturers of this institution, observing writing practices and interacting with the participants closely in the course of my work as a writing tutor, helped me 'build holistic understandings'. The use of a range of data types including interactions with subject teachers and course documentation have added to the 'talk around texts' taking this corpus linguistics study beyond the mere reporting of frequency counts.

The Civil Engineering Department at Middle East College was established in 2012. It offers two undergraduate programmes: *Civil Engineering* and *Quantity Surveying*. The duration of these programmes is four years across eight semesters of study. Each programme contains 25 modules and students are expected to take up to four modules in a semester (see Appendix 1 for the complete list of modules). The department has 19 faculty members from seven countries and the number of students enrolled on its programmes is approximately 450. None of the lecturers or students are L1 speakers of English. These participants fit my definition of 'expanding circles English as a lingua franca' group of countries where even though the medium of communication is English, the L1s of participants are not (see 1.1). More than half of the lecturers in the department have worked in the industry and all of them have a minimum of five years of teaching experience.

3.2. Data collection methods

This section sets out the details of how student assignments were collected to compile the corpus according to specific design criteria. It also describes how information from other sources of data, course documentation and specialist informants, informed the analyses of the texts and interpretation of the findings.

3.2.1. Corpus Compilation

To compile the corpus of Civil Engineering I collected student assignments from across all the semesters of study (except for Semester 8 assignments and one Semester 7 assignment which had not yet been submitted by students) in all but two of the modules in which written

work was assigned. The assignments were collected from both CE programmes, *Quantity Surveying* and *Civil Engineering,* since they share common modules (See Appendix 1) and because subject lecturers also mentioned that many common core concepts were also taught in both programmes. This justified including texts from both programmes in the same corpus. Henceforth this corpus will be referred to as the Omani Corpus of Academic Writing in Civil Engineering (OCAW-CE)). I have named it thus because I plan to extend this corpus to include other disciplines after the current project is completed. I collected assignments that received at least a B grade and are therefore considered proficient by the subject lecturers. Due to its means of collection and the linguistic context (see Section 1.1), the OCAW-CE can be characterized as a lingua franca corpus. Since one of the purposes of the study is to uncover the similarities and differences between genres, the choice of assignments which have been accepted by subject lecturers as fulfilling assignment expectations is appropriate in order to answer the research questions.

The collection processes for the BAWE (Alsop and Nesi 2009) and MICUSP projects (Römer and O'Donnell 2011) were reviewed as they too involve the systematic collection of large numbers of assessed student assignments. These projects, however, were large-scale in terms of the number of researchers, institutions, disciplines and assignments involved. Since, unlike these other projects, the current project was planned and undertaken by an individual, the data collection process was adapted to be less labour-intensive. For example, the BAWE and MICUSP corpus developers remunerated students for their contributions and installed manual and online systems for obtaining copyright and for authorial and textual metadata. Since I did not have the financial and other resources to follow these methods, I adopted a convenient and efficient system that would be appropriate for the resources at hand and that would suit the compilation of a smaller corpus. This process is described below.

First of all, the Head of the Department and faculty members of the CE Department were informed about the project through email and through informal discussions. All the lecturers interviewed were given Participant Information Sheets (attached to the Ethics Approval Request to Coventry University Ethics Committee) which provided them with details about the study. The Participant Information Sheet also informed lecturers of their

right not to be interviewed or participate. They were assured of confidentiality and the fact that the outcomes of the study would be shared with them. After obtaining approval from Coventry and Middle East College Ethics Committees to collect student assignments for research purposes, I informed the Moodle administrator about the project. Moodle is the virtual learning management tool used at the research site for lecturers to upload course documentation and for student assignment submissions. In line with the ethics application (reference P33371), a Consent Form was put up on the submission pages of Moodle explaining the purpose of the project and informing students of their right to withdraw their assignment from the study. The form assured students that their assignments would only be used for research purposes and that their identities would not be made public. Details such as my name and email (and that of my local supervisor) were provided so that students could contact us if they had any queries or if they did not want to participate in the study. In practice, none of the students objected to their assignments being used for research purposes.

The sampling method used in this study fits the definition of convenience sampling offered by Mackey and Gass (2012: 81):

> *Captive audiences such as students in the researcher's own institution are prime examples of convenience samples. To be fair, convenience samples are rarely completely convenience-based but are usually partially purposeful, which means that besides the relative ease of accessibility, participants also have to possess certain key characteristics that are related to the purpose of the investigation.*

I commenced with the data collection after all the steps above were completed. Every academic department at MEC has a CAW coordinator who is responsible for liaising with the CAW team on matters related to academic writing. To make the data collection process less labour intensive, the relevant CAW coordinator was requested to liaise with faculty members to collect assignment briefs (See Appendices 3A and 3B) and student assignments. This methodological procedure shortcut the process of contacting students individually and persuading them to contribute their assignments. The CAW Coordinator specified that assignments awarded an 'A' or 'B' grade usually exceeded the expectations of the subject

lecturers, and together we came to a common understanding that an assignment awarded an 'A' or 'B' grade would be considered 'proficient'. This information was conveyed to subject lecturers during a department meeting and they were requested to contribute three proficient assignments from each of the modules they taught. In fact it turned out that some teachers provided only one assignment sample while others supplied up to 18 samples from the same coursework. I decided to include these extra samples in my study as more texts would be useful for my purposes.

The method of data collection adopted for this corpus is much more convenient and efficient than that adopted by the BAWE and MICUSP project teams, as student assignments were gathered directly from the lecturers. I did not have to wait for volunteer students to contribute their assignments for my study and I could also ensure representativeness of my corpus by making sure that samples of assignments from the selected modules are included in the data. The process of obtaining student consent was also much easier through the Consent Form posted online. The metadata collection was done simultaneously with the data collection; lecturers provided the module name and code when the samples were supplied.

3.2.2. Processing of the Corpus

Metadata thus collected were transferred to a spreadsheet. The student contributors' identities were anonymised when metadata were recorded by removing the identifying information in the first page of the assignment. A unique id number was then assigned to each student and assignment. For example, if a student was assigned the id number CV0001 and the student had submitted more than one assignment, the samples would be named CV0001a, CV0001b and so on. This alphanumerical naming system was borrowed from the practice followed by the BAWE project team (a spreadsheet with BAWE assignment details is available from the Oxford Text Archive, and also from the BAWE webpage at Coventry University). Care was taken not to include more than six assignments submitted by the same student to prevent individual idiosyncrasies from distorting the results of the study. I am the only one who can trace the identity of the student as the files are stored on a private and password protected laptop. Even after the study is completed, the files will be stored in a

password protected location that can only be accessed by myself. This data will be helpful in my role in the institution as academic writing tutor and for future research projects.

The assignments were marked up using the xml text editor, *Oxygen,* which is compatible with the interface used for analysing the corpus, Sketch Engine. This text editor was used for marking up sections of the text based on their communicative function. The headings provided by the student and sometimes provided in the assignment briefs (e.g. Introduction, Methods) along with a close reading of the texts to identify the main functions of text segments allowed me to identify these sections. It is worth mentioning here that the headings provided by the student did not always reflect the content and therefore I had to read the assignments closely to identify the communicative intent of text sections. For example, in the Lab Reports, the results are sometimes reported in a section given the heading 'Discussion'.

The Introductions and Conclusions of assignments in particular helped in this respect, as did the assignment briefs. The interpretation process allowed me to annotate sections of the text such as 'Methodology' or 'Recommendations' or 'Literature Review' and involved multiple close readings of the text which resulted in decisions as to where to mark off the various sections of the text even if they were not explicitly identified by the student through headings or sub-headings.

The location of sections, headings, sub-headings, paragraphs, tables and figures was indicated by marking them up within angle brackets. Numerical information and special characters such as calculations and formulae were removed and replaced with code showing where they are located in the text. The idea of replacing calculations and formulae with codes was taken from the mark-up process followed by the BAWE project team (Ebeling and Heuboeck 2007). This mark-up process involved close analysis of the structure of individual texts.

3.2.3. Description of the corpus

Most researchers' definitions of a corpus uniformly make the case that just any collection of texts do not make a corpus (Rizzo 2010). The texts selected for inclusion in a corpus must satisfy the purpose for which the corpus is designed as well as fulfil criteria related to representativeness, balance and other features decided *a priori* or during the data collection process. Sinclair's (2005: 19) definition of a corpus emphasizes the aspects of corpus design that a corpus should fulfil. He defines a corpus as, 'a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research'. This section will attempt to describe the guiding principles behind the design of the Omani Corpus of Academic Writing-Civil Engineering (OCAW-CE) in terms of representativeness and balance to illustrate that it represents a language variety which can be used for linguistic research.

3.2.3.1. Representativeness, balance and size

Student assignments representing almost all the modules from the two programmes of Civil Engineering offered by the college are represented in OCAW-CE. This corpus contains samples from all the assignments with the exception of one which was required to be submitted in the form of slides, one consisting entirely of numerical data, and the coursework in Semester 8 which was not yet submitted by the students at the time of data collection (the programme had only been implemented until Semester 7). Hence this corpus is quite representative of the types of discursive assessments and the characteristics of Civil Engineering student writing in this institution.

The OCAW-CE is a synchronic corpus as texts were collected over the course of two semesters in the academic year 2014-15. This decision followed automatically after the first phase of data collection was completed as there were enough texts from two semesters to form a specialized corpus.

Regarding balance, I attempted to collect three assignments to represent each piece of coursework as the purpose of this corpus is to represent the writing that is required of a

student during the course of the Civil Engineering Programmes. As I did not receive the same number of samples for each coursework, the corpus is not fully balanced, although it contains samples of all of the coursework assignments and represents the assignments written by 59 students. Table 3.1 shows the number of assignments collected from each semester of study according to the Degree Plans of the two CE programmes.

Genre	No. of assignments	No. of words	Average no. of words	Semester
Case Study	31	105,806	3,413	1, 2, 4, 5, 6, 7
Methodology Recount	63	39748	641	1, 2, 3, 4, 5, 7
Explanation	20	68,041	3,402	1, 2, 3, 4, 5, 7
Exercise	28	28,692	1025	5, 6, 7
Site Investigation Report	4	9,577	2,394	3
Manual	3	8,390	2,797	4
Total	149	260, 453		

Table 3. 1. Number and distribution of texts across the semesters of study

With regards to the size of a corpus, as Sinclair (2004: 21) holds, 'There is no maximum size' by which he means the purpose of the study should guide the design of the corpus. This corpus comprises 260,254 words and 149 texts; OCAW-CE is relatively small in comparison with other corpora of student academic writing like the BAWE (about 6.5 million words) and MICUSP (about 2.5 million words). However, it should be noted that these corpora contain texts belonging to multiple disciplines. Viewed from this perspective, OCAW-CE is a substantial corpus of student writing in Civil Engineering (compared to 31 CE texts in MICUSP and 9 in BAWE). Further, as Sinclair argues (2005: 29), for specialized corpora 'a much smaller corpus will be needed for typical studies than is needed for a general view of the language'.

3.3. Other information sources

The corpus described above was complemented by other forms of data used in qualitative research: information collected via discussions with subject lecturers, and documentary evidence such as assignment briefs (see Appendices 3A and 3B) and course information guides. These qualitative data were used to guide the interpretation of the findings of the quantitative corpus approach. The conclusions drawn from quantitative results and analyses of the texts themselves were informed by the assignment briefs and the perspectives of subject lecturers. As examiners and assignment designers, it is important to find out what they value in student writing and how this influences their expectations about student assignments.

This integration of quantitative and qualitative methods in corpus studies is demonstrated by Nesi and Gardner (2012). Interviews with subject lecturers were used as a valuable source of data in their project. Nesi and Gardner (2006:102) describe the nature of undergraduate writing as 'complex, with many variations in practice dependent not only on discipline, level of study and educational approach, but also on the nature of the higher education institution, the particular focus of the department within that institution, and the idiosyncrasies of the lecturers who assign written work'. Therefore they made use of interviews with subject lecturers to identify 'important trends in the assignment of student writing tasks'.

Fortunately, as Director of the Writing Centre, I had routine interactions with the subject teachers as part of my job and I considered these discussions as part of an ongoing conversation.

I was therefore able to triangulate my information sources by considering discussions with subject lectures and course documentation along with the student assignments themselves to arrive at conclusions and thus increase the validity of the study.

These discussions, which were held over the entire course of this thesis and even before, explored these themes: what genres are preferred in the discipline, what lecturers

value in student writing, progressive difficulty of assignments over the semesters of study and student performance in them, the challenges they face, and suggestions for pedagogical interventions to improve the provision of academic writing support. The purpose behind the design of assignments, their genre labelling conventions and clarifications on their structure were additional areas explored.

3.3.1. Course documentation

Assignment briefs and 'Module Information Guides' provide information about the learning outcomes of the module, assessment types and deadlines, and other information such as the penalty for plagiarism. They are uploaded on the module Moodle page at the beginning of the semester and were also collected as part of documentary evidence to support the information gathered from the assignments. An examination of the briefs yielded useful information about the type of assignments designed, the value subject lecturers give to particular components of the assignment, student understanding of assignment expectations and other details which would contribute to the characterization of student texts.

In connection with documentary evidence, Merriam and Tisdell (2016: 175) point out that it 'evolves from the topic of inquiry itself.' Additionally, 'most are produced independently of the research study. They are thus nonreactive and grounded in the context under study'. Studies that aim to classify student writing at university use assessment briefs as a means of identifying and categorising assignment types (see 2.4 for a detailed review of these studies). Similarly, this study used assessment briefs as an additional source of data to gain insights not only into the types but also the characteristics of student writing which may have been influenced by the design of the assignment. The authenticity and accuracy criterion that is used to verify documentary evidence is easily met with this data source, as these documents are pre-existing and stored in a central repository (staff documents server) and are retrieved with the permission of the faculty members. I was also well aware of the caveat that assignment briefs on their own may not provide enough information for classification purposes (Gardner and Nesi 2012). However, taken together with student assignments and

discussions with faculty, these documents informed the genre labelling process and helped to provide a holistic picture of disciplinary writing requirements.

3.3.2. Specialist Informants

The specialist informants consulted during the course of this study are the lecturers from the Civil Engineering department. In my study, these informants fit the description of 'key informants' provided by Merriam and Tisdell (2016: 129).

> Key informants are able, to some extent, to adopt the stance of the investigator, thus becoming a valuable guide in unfamiliar territory. But not all good respondents can be considered key informants in the sense that anthropologists use the term. Good respondents are those who can express thoughts, feelings, opinions—that is, offer a perspective—on the topic being studied. Participants usually enjoy sharing their expertise with an interested and sympathetic listener. For some, it is also an opportunity to clarify their own thoughts and experiences.

Appointments were made with ten subject lecturers from the Civil Engineering Department at MEC for discussions about the student assignments. All the lecturers had been working in the college for more than three years and taught all levels of the programme. They were given Participant Information Sheets (submitted along with the Ethics Application to the CU Ethics Committee) in which they were told about the purpose of the study; they were also reassured that their identities would be kept confidential if they decided to participate in the discussion. They did not mind the conversation being recorded and also gave permission for their words to be quoted in my thesis. A sample transcript of one of these conversations is provided in Appendix 4. It should be borne in mind that I routinely interacted with these lecturers in the course of my work at the Writing Centre; therefore insights obtained during these other interactions also informed some of the conclusions and interpretations I brought to my study. It was, however, highly unlikely that my position as Director of the Writing Centre had any effect on the opinions expressed by the subject lecturers, since the subject lecturers were rather territorial regarding their disciplines, and would not allow an outsider to influence

their expectations of the assignments they set, or their understanding of the discourse of their fields.

The CE lecturers helped me navigate the sometimes 'unfamiliar territory' of student assignments including their purpose and structure. They were also able to explain their expectations of student writing in line with disciplinary and department requirements. Most of my specialist respondents were empathetic to the purpose of my study because they felt that ultimately it would help the students write better. They also sometimes reflected on the design of their own assignments and spoke about pedagogical issues that arose during my routine interactions with them as part of my work at the Writing Centre. For example, when a particular issue was raised such as the obligatory inclusion of recommendations in Case Studies or use of Location statements when a figure is included, subject lecturers voiced their opinion that perhaps there should be more explicit teaching. Thus, including the viewpoints and experience of subject lecturers has complemented the use of the other data sources in my study (student assignments and course documentation). Denzin's (1978) (cited in Merriam and Tisdell 2016) discussion of triangulation involves four types: multiple methods, multiple sources of data, multiple investigators and multiple theories. The validity of this study is strengthened by triangulating methods of data collection through student assignments, assignment rubrics and discussions with specialist informants.

3.4. Categorization of texts into genres

The assignments in OCAW-CE were categorized into genres on the basis of Nesi and Gardner's (2012) genre classification framework (see Section 2.2.4). This process addresses one of the main requirements for answering the research questions of this study which involved uncovering the similarities and differences between the two main genres of CE.

When I examined the assignment labels provided by the subject lecturers and the assignments, I found that there were mismatches as these labels did not always reflect the purpose of the assignment. For example, an assignment which contains mainly explanations might be called a Case Study. As Nesi and Gardner (2012: 5) point out, although 'the

nomenclature used within university departments to specify different assignment types goes some way towards identifying and distinguishing those genres', a more robust classification system of the writing genres I encountered was required. I followed the approach adopted by the BAWE and MICUSP project teams; they generated a data-driven taxonomy of writing genres of university students when they recognized the mismatch between the department and student labelling of assignments and the actual student assignments.

I decided to adopt Nesi and Gardner's (2012) framework for several reasons. An examination of these two systems of classification revealed that in the MICUSP framework, characteristics of assignments such as their evaluative, creative or argumentative quality were recognized as 'text types'. These are not genres and therefore the MICUSP taxonomy is not suitable as a framework to categorise genres. For instance, it would be difficult to classify assignment types like Case Studies and Site Investigation Reports which I found in my Civil Engineering corpus. The taxonomy developed by Nesi and Gardner, on the other hand, was based on the primary communicative intent of the assignments; they term this the 'social function' of these assignments. This taxonomy was suitable for my purposes as it would accommodate the categorization of any type of assignment, being based on their broad communicative purpose rather than specific characteristics of texts. Their examination of the rhetorical sections of texts along with the exemplars provided was also a useful guide during the genre categorization process.

3.4.1. Procedure for genre analysis

The first step in the classification process was to identify the genre of texts by examining the headings and sub-headings provided by the student to help me identify the purpose of a text. Documentation consisting of assignment briefs (see Appendix 3), notes taken during the mark-up process, and periodic discussions with specialist informants (subject lecturers in Civil Engineering at the research site) also greatly contributed to the classification process. Identifying the main purpose and the rhetorical sections enabled me to map each text to Nesi and Gardner's (2012) genre classification system. When classifying each text, I first read through the assignment brief to understand the purpose of the assignment and to identify

other clues provided as to the nature of the writing task, such as the labelling of the assignment, and instructions to students on the staging of their texts and the content of each section. The student assignments were then read closely with attention to staging and conformity to the instructions in the assignment brief. It was at this point of time that subject lecturers were consulted and asked to explain the purposes of the assignments.

For example, the following is a statement found in one assignment brief:

> *The purpose of this coursework is to understand to what extent local sustainable projects in Oman are actually performing up to the standards defined in the scientific literature with regards to sustainable development in construction and built environment. To do so literature review must be conducted to be able to first understand the principles of sustainability. Moreover a model should be prepared to define appropriate factors and criteria in order to be able to measure the performance of the case-study*.

I was able to deduce from this assignment brief statement, and from the student assignment itself, that students had to first review the literature, identify a model of assessment in order to assess a 'case', analyse the case, and then provide recommendations. I came to the preliminary conclusion that this assignment belonged to the Case Study genre family, as described in Nesi and Gardner (2012). I then met with the subject lecturer for this module and he explained that this task enabled students to practise the type of writing expected of practising engineers.

Nesi and Gardner (2012: 189) explain that in BAWE, 'the Case Study genre family contains texts which analyse an exemplar in order to demonstrate or develop an understanding of professional practice. They always include recommendations or suggestions for future action.' The assignment discussed in the previous paragraph required students to identify a single 'green building' in Oman and then assess its performance in terms of sustainable and 'green construction' practices. The aim of the task was to prepare them for professional practice, and it clearly required the students to 'analyse an exemplar', and provide recommendations.

This was the procedure that I followed to categorise all the assignments in my corpus according to the 13 genre families described by Nesi and Gardner (2012).

Sub-corpora of texts belonging to a genre were then created by grouping texts belonging to the same genre in separate folders and uploading them to *Sketch Engine* (see Section 3.5 for details of *Sketch Engine*). This was done to address the questions of this study which are to identify the two main genres in CE and describe them linguistically in terms of phraseology and phrasal complexity.

3.4.2. Inter-rater reliability

The OCAW-CE corpus comprises 260,453 words and contains 149 student assignments collected from across semesters of study. To achieve consistency and reliability in the categorization of assignments in the corpus into genres, measures were taken to ensure inter-rater reliability among three raters. Inter-rater reliability, in the context of this study, can be defined as the degree of alignment of categorization decisions among the three raters when texts are classified independently. Testing for inter-rater reliability is necessary since 'data are reliable if coders can be shown to agree on the categories assigned to units to an extent determined by the purposes of the study' (Artstein and Poesio 2008: 557). If the raters make consistent decisions based on pre-determined coding guidelines, the procedure can be expected to produce similar interpretations. On the other hand, a low agreement between raters indicates that more training on categorization is required or the coding guidelines may need to be clearer.

There were three raters involved in the categorization of the texts in the OCAW-CE into genres using the framework developed by Nesi and Gardner (2012). The first rater (myself) has many years of experience in the Writing Centre guiding students with their assignments. The second rater is the first of the two researchers who developed the classification framework (Nesi and Gardner 2012) and therefore had a good understanding of the bases on which the genre classification system was created. She has also had the experience of having rated thousands of similar student texts. The third rater who has more

than 15 years of experience in EAP was able to contribute so that an unbiased agreement on the guidelines could be reached based on discussions of individual texts.

The inter-rater reliability testing and the procedure for classification followed in this study has been partially adapted from Fuoli and Hommerberg (2015) and Moreno and Swales (2018).

3.4.2.1. Devising guidelines for the classification scheme

I went through about 35% of the assignments in the corpus to gain an understanding of text categories and map them against the genre classification framework. This sample was chosen to represent, as much as I could discern, all the types of assignments in OCAW-CE. The first step of this initial categorization process involved segmenting all the texts into rhetorical stages to devise the first draft of a set of guidelines after consulting the course documentation and subject lecturers. A month later, the texts were categorised by the first rater once again to check for human errors and internal consistency and to refine the guidelines as reported by Moreno and Swales (2018).

3.4.2.2. Training of raters

The three raters then independently categorised these selected assignments. Disagreements were reconciled and further refining of the guidelines for categorizing the remaining texts was done. After this step, eight to nine assignments at a time were examined by the three raters over six rounds, followed by discussion sessions after each round. Where there was disagreement, the discussions involved going back to the initial understanding of the framework and segmenting the text into rhetorical stages once again to uncover the genre. This also sometimes entailed further discussions with the subject lecturer and reading the course documentation to check the expectations of the subject lecturer and the objectives of the task. I also took notes for the individual assignments discussed as reference in the event of similar assignments in the rest of the corpus. Care was taken to include different types of assignments so that the remainder of the texts could be interpreted based on these discussions among the three raters.

3.4.2.3. Inter-rater reliability test

The percentage agreement test was adopted to measure inter-rater reliability in this analysis. It is a simple statistical test to report inter-rater reliability showing 'the number of agreements per total number of coding decisions' (Biber et al. 2007). Since about 50% of the texts were examined by all three raters and all the text categories emerging from the data were examined and also because the categorization process involved only one level of granularity, other tests like the Cohen's Kappa coefficient which allows for a chance-corrected agreement were not applied. This test was mentioned by Moreno and Swales (2018) in their description of the inter-rater reliability tests they performed to check the consistency of coding decisions during the segmentation of research articles into moves and steps. Of the 75 texts rated by all the three raters independently, there was an agreement percentage of about 81%. After discussion, the agreement percentage reached 100%.

I then categorised the rest of the texts according to the guidelines and the notes taken for each assignment during the training phase. If there were any doubts at this stage, a discussion session was held to reach an agreement. Disagreements arose mainly due to rater fatigue and distraction and were soon resolved. The entire process including all the four stages took about four months to complete.

3.5. Corpus analysis tools

I used two corpus tools, AntConc (https://www.laurenceanthony.net/software) and Sketch Engine (http://www.sketchengine.co.uk/), to conduct the analyses required to answer the three research questions of my study.

3.5.1. AntConc Concordancer

I used the AntConc tool for retrieving Lexical Bundles (LBs) from the Case Study and Methodology Recount sub-corpora. I needed to retrieve the LBs used in student assignments to answer Research Question 2 on the differences and similarities in phraseology between CS and MR assignments. I found that although Sketch Engine allowed me to set frequency thresholds, there was no option to set dispersion cut-offs. Since AntConc offers the facility

of setting both frequency and dispersion thresholds while retrieving LBs, I chose to use this corpus tool to retrieve the LBs in the Case Study and Methodology Recount assignments.

3.5.2. Sketch Engine

Sketch Engine was one of the corpus tools used in this study to examine the phrasal complexity features and *the N of (the) N* patterns in the CS and MR sub-corpora as it offers the features I required to conduct the analyses. I also used it to share my corpora with the supervisors of my study. Although Sketch Engine accepts various text formats, the texts in OCAW-CE were converted to xml format because I had marked up the rhetorical stages and other linguistic features in each text such as visual data and mathematical expressions. I saved the assignments in separate folders according to their genre categories and the files in each sub-corpus were then zipped into a single file and uploaded on Sketch Engine (CS, MR, Explanation, Exercise, Site Investigation Report, Manual) separately; hence, I divided the OCAW-CE corpus into six sub-corpora. I chose the option 'Expand this archive instead of converting it to a single plain text' each time I uploaded a file because I wanted the names of individual files to be listed when the concordance lines are displayed. The next step involved the automatic corpus compilation stage. I chose the recommended default option *English 3.3 for TreeTagger pipeline v2* tagset (a set of part-of-speech tags which labels each token in a corpus) to automatically tag all the texts.

My Research Question 3 on the differences in phrasal complexity between CS and MR texts required the generation of nominalisations, noun strings and attributive adjectives and also the retrieval of a string of words having *the N of (the) N* pattern. Corpus Query Language (CQL) is a feature of Sketch Engine that I used to retrieve these words and strings. This query language facilitates advanced searches including the retrieval of strings that have certain patterns and parts-of-speech, made possible because the corpora uploaded on Sketch Engine are tagged for parts-of-speech when they are uploaded (see 3.7 for more details).

Both AntConc and Sketch Engine displayed the results generated by these searches in the form of lists and KWIC format. In the KWIC format, the search word or phrase (node) is highlighted in the centre with some context to the left and right of the node. I could also

sort the words on either side of the node and read the expanded context for qualitative analysis.

3.6. Analysis of Lexical Bundles in the MR and CS sub-corpora

The study of Lexical Bundles (LB) is a sub-branch of phraseological research. I decided to focus on LBs in my study of the phraseology of the MR and CS assignments since it is a well-established research tradition compared to p-frame work and also, the methodological procedures tend to be more comprehensive (see Section 2.7). LB studies have mainly focused on comparisons between disciplines and groups of writers (see 2. 5). However, studies of LBs across genres in specific disciplines, especially in undergraduate writing by students is a less explored area. I addressed this gap by conducting a quantitative and qualitative study of LBs in the two major genres (Case Studies and Lab Reports) in the Omani Corpus of Academic Writing-Civil Engineering (OCAW-CE). This study attempted to answer Research Question 2.

Section 3.6.1 will discuss the procedure used to generate LBs from the two sub-corpora. Section 3.6.2 will discuss the structural and functional categorization of LBs using the frameworks developed by Biber et al. (1999) and Hyland (2008).

3.6.1. Corpus analysis techniques to retrieve LBs in CS and MR assignments

I used the AntConc program to generate 4-word bundles from the Case Study and Methodology Recount sub-corpora. I set different dispersion cut-offs to retrieve LBs from the CS and MR sub-corpora since the number of texts and the size of the two sub-corpora differ.

I decided to focus on four-word bundles for my study because the literature (Hyland 2008) indicates that they are much more frequent than five-word bundles and also that they typically have clearer structures and functions than three-word bundles. It will also be useful to compare the findings of this study with many other studies that have chosen to examine four-word LBs. The phenomenon of ‘nesting’ as explained by Biber et al. (1999) demonstrates how four-word bundles are frequently nested inside longer bundles. Therefore,

if a particular four-word LB type is found to be highly frequent, deeper investigations can follow; such LBs can be extended to include more words preceding and following them (see 3.7.2 for an explanation of how a frequent LB type *the N of the* was extended to the pattern *the N of (the) N*).

The cut-off rates used in other studies are rather arbitrary and not always clearly justified. A frequency threshold of 40 pmw was used by Biber et al. (2004) and Biber and Barbieri (2007) but Hyland (2008) used a lower cut-off of 20 per million words (pmw). With regard to dispersion of LBs across texts in a corpus, most studies have adopted a dispersion rate of occurrence in at least 3 to 5 texts (Biber and Barbieri 2007, Biber et al. 2004, Chen and Baker 2010, Cortes 2004). Hyland (2008) set the dispersion criterion as 10% of the total number of texts while Rezoug and Vincent (2018) set it at 20%.

Dispersion rates were set at five for the MR sub-corpus and three for the CS following Pan et al.'s method of setting dispersion thresholds; they used equal sized corpora but chose different dispersion rates because the number of texts in their corpora varied. With regard to frequency thresholds, I followed Bestgen's (2019: 15) recommendation of a 'dynamic threshold for frequency' to avoid the effects of the Zipfian Law which stipulates that 'overrepresentation of rare items is larger for small texts and corpora' (12). This proposed solution (though 'arbitrary') consists of applying different frequencies according to corpus size (e.g. Chen and Baker 2016) although in my case, I am applying a higher dispersion cut off for a smaller corpus to reduce the effects of the Zipfian law.

The dispersion threshold also had to be adjusted while setting the frequency because frequency (raw count) cannot be less than dispersion. At first, I thought about increasing the dispersion threshold of MR texts to 6, as there are approximately double the number of MR texts (63) compared to CS assignments (31). But this would also require increasing the frequency to 6, in which case only 32 LB types would be generated, as against 111 LB types generated from the CS corpus when a frequency of 5 and dispersion threshold of 3 was applied. The frequency and dispersion were thus adjusted following Bestgen's (2019) advice so that the number of LB types generated from the two sub-corpora was somewhat similar

(see Table 3.2). Chen and Baker (2010) and Cai (2016) also speak about 'repeated experiments' to determine the thresholds. Applying the resulting thresholds resulted in the generation of LBs with frequency thresholds of 40 pmw and 102 pmw in the CS and MR sub-corpora, respectively. Table 3.2 contains the thresholds and the resulting number of LB types that were generated from each of the sub-corpora.

	Frequency	**Distribution**	**Min. freq. threshold (pmw)**	**Number of LB types**
MR	5	5	102	126 LB types
CS	5	3	40	111 LB types

Table 3. 2. Frequency and dispersion threshold for extracting Lexical Bundles

While processing the texts in the corpus and converting them to xml format, I had removed mathematical formulae and expressions and replaced them with <formula> thus indicating their location. This step ensured that mathematical characters were not retrieved as part of LBs. They were excluded from the corpus because my study is focused on the linguistic features of texts; however, I needed to mark their location because I was planning to study the organization of assignments in terms of moves and steps.

3.6.2. Structural and functional classification of LBs

Once retrieved, the LBs in the OCAW sub-corpora of Case Studies and Methodology Recounts (also referred to as Lab Reports) were categorized structurally, according to Biber et al.'s (1999) classification scheme, and functionally, based on Hyland's (2008) classification scheme (see Section 2.7.5).. These taxonomies were explained in 2.7.3 and 2.7.5. Biber et al. (1999) identified 9 structural patterns of LBs in academic prose (and also a category labelled 'other' for miscellaneous LBs that did not match any of these 9 patterns), of which Hyland found only 7 patterns. The procedure that I followed to classify the LBs according to their structure in the OCAW sub-corpora was fairly straightforward since I

could look at the examples provided in Biber et al. (1999), Hyland (2008) and Rezoug and Vincent (2018).

The functional classification scheme designed by Hyland (2008) was more difficult to apply. This difficulty of applying functional frameworks has been discussed in other studies. Adel and Erman (2012) point out that Biber et al.'s (2004) scheme provides no clear indication as to which sub-category a particular bundle should belong to; for example, the Focusing sub-category is named Discourse Organising in Chen and Baker (2010) while Simpson-Vlach and Ellis (2012) and Biber et al. (2004) categorise it under the Referential category. With regard to Hyland's functional framework (2008), Rezoug and Vincent (2018) point out that while the main functional categories may be clear, the sub-functions are at times ambiguous.

Neither Hyland (2008) nor Biber et al. (2004) explain how they dealt with multi-functional bundles; for example, the LB *at the same time* could convey the meaning of simultaneity or function as a transition, meaning 'on the other hand'. They also do not provide the complete list of LBs in their corpora under the various sub-categories, making it challenging for other researchers to replicate the procedures followed in these studies.

This ambiguity in the methodology has been addressed to some extent by Rezoug and Vincent (2018) who have provided comprehensive lists of LBs in each of the four sub-corpora in their corpus. It was thus possible for me to refer to their list to look at examples if I was faced with any doubt regarding how to classify any LB according to its sub-category, all the time bearing in mind its use in my own corpus. Rezoug and Vincent also explain their methodology in detail, unlike Hyland (2008). In the case of multi-functional bundles, they read the expanded context of the LB to identify its most frequent function. This was the procedure that I followed if there was an indication that a particular LB could have more than one function.

As in other studies (Hyland 2008, Rezoug and Vincent 2018) I also investigated the links between the structural patterns of LBs and their discourse functions since these previous studies have established clear associations between the two.

There is ample evidence in the literature which demonstrates the usefulness of studying LBs. In spite of not being perceptually salient, they serve clear functional purposes and studies have shown that LBs are a pervasive phenomenon in both spoken and written language (Biber and Barbieri 2007). Researchers also acknowledge disciplinary (Durrant 2017, Ebeling and Hasselgård 2015) and genre variation (Durrant 2017) in the use of LBs. As Biber and Barbieri (2007: 265) claim, they are 'not an accidental by-product of corpus frequency analysis' and therefore merit research attention. Since the purpose of my study is to uncover similarities and differences between the genres of Methodology Recounts and Case Studies, examining the LBs in them will offer useful insights into the phraseology of the two genres.

3.7. Corpus analysis techniques for extracting and investigating features that contribute to text density

3.7.1. Extracting noun strings, nominalisations and attributive adjectives in CS and MR assignments

Noun strings and nominalisations have been recognized as an important feature of academic writing, with advanced writers using it for various communicative rhetorical purposes (see 2.8.1). I will examine the use of these features in two of the genres that I have chosen to study extensively in this thesis: the Case Study and the Lab Report genres. There are two primary reasons why I chose to examine these features in my corpus. Firstly, their prominent role in academic writing is acknowledged by linguists (See 2.8.1, 2.8.2), and secondly, findings that have indicated that the occurrence of these phrasal complexity features in academic writing may vary according to discipline and genre. A description of their frequency and purpose in a specific genre and discipline in OCAW-CE will contribute towards the characterization of the texts in my corpus, the main aim of my study. These analyses will also have some pedagogical implications.

Sketch Engine allows the use of Corpus Query Language (CQL), which provides an easy and quick method of retrieval of nominalisations, noun strings and strings of words whose POS can be pre-specified in the query. CQL is a query language which can be used with the Sketch Engine interface to retrieve complex grammatical or lexical patterns which

cannot be otherwise retrieved using the standard user interface options available. The method of retrieval of noun strings using CQL is also fairly straightforward with Sketch Engine. This process is explained in this section.

3.7.1.1. Retrieval of nominalisations in Case Study and MR texts (OCAW-CE and BAWE sub-corpora)

The choice of nominalisations for analysis is based on Biber's (1988) identification of nominalisations. He considers nominalisations to be derived forms of verbs and adjectives, and identified as a nominalization 'any noun and its plural forms ending in *-tion, -ment, -ness,* or *-ity* (See 2. 6.1). Although this classification scheme does not identify all instances of nominalisations in the corpus, I decided to adopt this scheme to facilitate comparisons with many other studies that have used MDA or MAT software (Gardner, Nesi and Biber 2018, Nini et al. 2017, Staples et al. 2016).

Corpus Query Language (CQL)

Initially, I used the following CQL query to retrieve all words ending with the selected suffixes: [lemma=".*ity|.*tion|.*ment|.*ness" & tag="NN|NNS"]

This query retrieves all words ending with the 4 'Biber suffixes' which are also tagged as nouns (not including proper nouns; see the complete part-of-speech tagset[1] used by Sketch Engine).

As with all searches of this type, a precision check was needed to discover the extent to which I was retrieving nominalisations accurately, so I also manually checked all the concordance lines generated by the query.

The noun *department* (Example 01) is not a nominalization because it is not derived from a verb or adjective denoting a process or a quality (it is not related to the verb *depart*; see Section 2.8.1); this is a congruent and non-metaphorical use of language and not an instance of grammatical metaphor.

[1] The tagset for Sketch Engine can be found here: https://www.sketchengine.eu/english-treetagger-pipeline-2/

(01) There is a delay of payment for the work done because of the bureaucracy in government ***departments***.

I manually checked every concordance line to identify words such as this which had been wrongly retrieved as nominalisations; examples include *city, universities, mention,* and *business*. These items were then included in the CQL query to specify that they should not be retrieved.

Other words which caused some difficulties were those with several senses. One example is *management,* which could mean a group of people managing a company, the process of managing people or events, or even the name of a discipline (see Table 3.3). It can only be considered a nominalization according to the operational definition adopted in this study when it refers to a process or quality. I looked at the concordance lines and found many instances where it has been used to denote a process (Concordance Lines 1 to 4) but also a few instances where it refers to the group of people managing a company or organization (Concordance Line 5). In instances where it refers to a process, I considered *management* as a nominalization that would be included in my list.

1	activity of any successful organization is the	**management**	of this valuable asset. This knowledge
2	of this valuable asset. This knowledge	**management**	process consists of four main elements:
3	The knowledge in marketing, supply chain	**management**	, web site maintenance and other
4	This is the most essential task of knowledge	**management**	. Knowledge is worthless unless it is
5	that factors of risk and possible changes in	**management**	and workforce should be taken into

Table 3. 3. Concordance Lines with management as the node word from the BAWE Sub-corpus of Engineering Case Study assignments

With all the nominalisations included in the list, I tried to identify the verb and adjective forms from which they were derived. For instance, the use of the word *Information* (in the CS sub-corpus) in Example 2 is not an instance of nominalization because it is a part of a proper noun string referring to the name of a field of study and therefore was not included in my final list of nominalisations.

(02) …last point will be describe the importance of Information Technology (IT) applications in Quantity…

After conducting similar analyses for words identified as nominalisations following concordance analysis, the following operational definition for nominalisations was used in this study.

In my study, nominalisations are abstract nouns ending with the suffixes *-ity, -ment, -ness* and *-tion* which have been derived from verbs or adjectives and have some semantic correspondence with these root forms, or in the case of root forms having several meanings, the derived forms should correspond semantically with one of the meanings of the root forms.

The revised CQL query used to retrieve nominalisations from the OCAW-CE Case Study sub-corpus is given below. The exclamation mark (!) is used within the CQL to exclude specific words such as *city* and *university* which were wrongly retrieved because of their endings.

[lemma =".*ity|.*tion|.*ment|.*ness" & tag="NN|NNS" & lemma!="city|university|entity|authority|element|equipment|abutment|business|basement|catchment|cement|cavity|corporation|condition|capillarity|department|cross-section|cement|document|embankment|federation|fraction|foundation|friction|function|gravity|station|pigment|environment|monument|nation|institution|instrument|ration|segment|two-section|supplement|amenities|commodities|community|intercity|Segment|section|sediment|traction|government|question|priority|portion|quality"]

Limitations of the method

In Examples 03 and 04, the nominalization *emissions* is derived from the verb *to emit* to refer to a process, and *efficiency* from the adjective *efficient.*

It can be seen that such nominalisations will not be captured if only the endings considered by Biber (1988) are taken into account. I decided to work with this limitation, however, because this method helps retain comparability with earlier research.

(03) Therefore, Middle East countries are known for high ***emissions*** *of carbon dioxide per capita*

(04) Further to the above, there are several factors that influence in creating green building which are energy ***efficiency****…*

3.7.1.2. Retrieval of Noun Strings and of Attributive Adjectives

Several studies (Gardner, Nesi and Biber 2018, Staples et al. 2016) have indicated that scientific writing is characterized by the use of noun strings. Their ability to compress information and express scientific concepts has also been well-documented (Biber and Gray 2016). I retrieved strings containing up to four noun modifiers to find out the proportion of noun strings of different lengths found in the two sub-corpora and thereby find out the extent to which the writers of CS and MR assignments exploit this device. This analysis can also contribute to the discussion in the literature about the low number of noun strings containing more than one or two modifiers (Montero 1996). To this end, the normalized frequencies of noun strings in the CS and MR sub-corpora were compared. The types of pre-modifiers used were then examined to investigate the meaning relationships between the pre-modifiers and the Head Nouns. Examining these meaning relationships revealed the various functions realised by these strings. For example, some pre-modifiers (e.g. *wood, rock*) were Concrete nouns used to specify the material that the Head nouns were made of (e.g. *chunks, samples*). This analysis was motivated by Biber and Gray's (2016) analysis of the noun strings they found in their corpus. Table 3.4 contains the CQL queries used to retrieve noun strings of different lengths.

No. of pre-modifiers	**CQL**
One pre-modifier	[tag="NN|NNS|NP|NPS"][tag="NN|NNS|NP|NPS"]
Two pre-modifiers	[tag="NN|NNS|NP|NPS"] [tag="NN|NNS|NP|NPS"][tag="NN|NNS|NP|NPS"]
Three pre-modifiers	[tag="NN|NNS|NP|NPS"][tag="NN|NNS|NP|NPS"][tag="NN|NNS|NP|NPS"][tag="NN|NNS|NP|NPS"]

Four pre-modifiers	[tag="NN|NNS|NP|NPS"] [tag="NN|NNS|NP|NPS"][tag="NN|NNS|NP|NPS"][tag="NN|NNS|NP|NPS"][tag="NN|NNS|NP|NPS"]

Table 3. 4. CQL used to retrieve noun strings of different lengths from CS and MR assignments

As an important aspect of noun phrase complexity (Biber, Gray and Poonpon 2011, Parkinson and Musgrave 2014), it seemed useful at this stage also to investigate the use of attributive adjectives. I therefore checked the frequency of attributive adjectives used before nouns retrieved to check whether there were any differences in their use in the CS and MR sub-corpora. Because of the design of my corpus (two genres from the same discipline written by students at approximately the same level of language proficiency and sometimes by the same student), I could investigate if genre can be the influencing factor for the choice of attributive adjectives by comparing the frequency of this linguistic feature in the CS and MR assignments. If the frequencies are different, I can conclude that the use of attributive adjectives depends on genre demands. To retrieve the attributive adjectives placed before nouns in the two sub-corpora, I used the following CQL query: [tag="JJ"] [tag="NN|NNS|NP|NPS"]. Using this query resulted in attributive adjectives being followed by nouns that were sometimes head nouns (e.g. *potable water*) and sometimes noun pre-modifiers of head nouns (e.g. *traditional water source*).

3.7.2. Extracting the N of (the) N pattern from the CS and MR sub-corpora

Another strand of this thesis relates to phrasal complexity in the CS and MR genres. In order to investigate phrasal complexity in the two genres (Research Question 3), I first retrieved noun strings with the pattern, *the N of (the) N* (as indicated below). This allowed me to uncover the similarities and differences in this aspect of the phraseology of the genres. This pattern was chosen because the *noun phrase with of-phrase fragment* was identified as one of the two most frequent structural patterns in both the genres in the Lexical Bundle (LB) analysis (see Chapter 5) but this allowed me to overcome some of the limitations of the LB approach and view this phenomenon from a different perspective. The pattern was extended

to include the second head noun since initial analysis indicated that the nouns used in the two genres differed, suggesting that examination of the second head noun would reveal further aspects about realisations of this pattern, such as preferred noun category sequences (see Section 3.7.2.1). Two analyses were then conducted for the purpose of uncovering the similarities and differences between the two sub-corpora in terms of phrasal complexity: 1) categorisation of the noun types constituting the first and second head nouns within *the N of (the) N* pattern and 2) comparison of the proportion of the various noun category sequences within *the N of (the) N* pattern across the two genres.

I used the following query to generate all instances of *the N of (the) N* pattern from the two sub-corpora (with the default attribute set as *lc=word (lowercase)* to ignore case during the search):

```
"the" [tag="N.*"] "of" [word="the"]? []{0,2} [tag="N.*"]
```

This query enabled me to generate instances of this pattern even if the second head noun occurred up to two spaces after the second optional 'the'. Adding two optional words after the optional 'the' was found to be optimal because this query was usually able to capture instances of the second head noun within the node group or close to it. The precision rates were very high (almost 100% if duplications are not considered) as almost all the concordances retrieved fit the pattern; the recall was enhanced by adding the two optional spaces before the second noun (from the CS sub-corpus, 2,141 instances were retrieved with two spaces while only 1,070 were retrieved without spaces).

Although the precision rate was high, I found considerable duplication in the concordances generated by this means, and had to manually eliminate these. The concordance lines (Examples (05) and (06) illustrate the duplications that had to be manually eliminated. The same line is repeated; (05) and (06) are both retrieved because the same sentence realises the pattern *the N of (the) N;* in (1) *seawater* is considered the second noun, while in (2), *desalination* is considered the second noun. The duplication occurred because the query retrieved each instance of *the N of (the) N* pattern irrespective of whether they occurred in a line previously retrieved.

(05) …system. Untraditional methods: including ***the process of seawater*** *desalination and wastewater treatment to make*

(06) …system. Untraditional methods: including ***the process of seawater desalination*** *and wastewater treatment to make the water*

After the duplications were discarded, there were 1314 instances of this pattern in the CS sub-corpus and 586 instances in the MR sub-corpus. These were then filtered down to 1280 lines in the CS sub-corpus and 583 lines in the MR sub-corpus; some of the remaining concordance lines were eliminated from the final list because they did not fit *the N of (the) N* pattern; on closer inspection, the second item tagged as a noun belonged to other word classes or the clauses did not convey any meaning. In concordance line (07), the second head noun is an adjective while in concordance (08), the clause does not make sense.

(07) …to form the rust. On the surface of the iron ***the crust of reddish brown*** *is formed. The iron is used so widely, for…*

(08) …in the future, and 32,000 square meters and ***the wishes of the eggs*** *to be sustainable and environmentally*

The next step was to identify the first and second nouns in the pattern; identifying the first noun was fairly easy since it invariably occurred in the second position of the node group of words. The second noun, however, posed a slight challenge since the head noun was not always in the same position; there were sometimes two or three pre-modifiers before the second head noun (**bolded**) and they sometimes occurred outside the node group, as Examples (09) and (10) illustrate:

	Node Group	
(09) a portion of the previously mixed soil into	*the cup of the liquid limit*	***apparatus*** *at the point where the cup rests on the*
(10) back to front by using grooving tool. Rotate	*the handle of liquid limit*	*testing* ***device*** *to begin vibration and counting*

Table 3. 5. Examples of concordances where the head noun is outside the node term

Since the categorisation process required manual analysis, the concordance lines were added to a spreadsheet to facilitate this. This process is explained in the next section.

3.7.2.1. Categorisation of noun types used by MR and CS texts

I categorized the nouns in the *Noun of (the) Noun* strings into ten main types, using a framework adapted from Biber (2006), with two additional categories: *gerund* and *Proper Noun.* The framework used to classify the nouns in *the N of (the) N* pattern in the two sub-corpora is given in Table 3.6.

Table 3. 6. Noun category framework for classifying nouns in CS and MR assignments

3.7.2.2. Biber's (2006) noun category framework

Biber's (2006) noun category framework used to classify the noun types in the T2K-SWAL corpus consists of eight main noun categories. Table 3.7 provides examples of nouns from the T2K-SWAL corpus classified according to this framework.

Table 3. 7. Semantic noun categories framework (Biber 2006)

Biber (2006) provides the complete list of nouns by category found in the T2K-SWAL corpus. I found that my noun list did not have much in common with this list; there was an

overlap of only about 17% of the nouns. This can mainly be attributed to the differences in topic and composition of the two corpora; the 2.7 million-word T2K-SWAL corpus contains only a small proportion of texts related to engineering, and moreover, comprises spoken and written texts used in university contexts such as course packs, classroom management, textbooks, and service encounters. Unlike the OCAW-CE, this corpus does not include any student coursework.

The noun list provided by Biber (2006) was thus used as a rough guide to categorise the nouns in my sub-corpora after consideration of their context, but I needed to use my judgement in most cases. There were some nouns in the T2K-SWAL which were similar to the nouns I found and helped with my categorisation process; for example, Biber (2006) places *thesis* in the *Technical/Concrete* category and so I found it easier to place the noun *project* which I found in my sub-corpora in the same category.

Another problem with categorizing nouns solely on the basis of the list provided by Biber (2006) resulted from the various semantic categories each noun can realise. Biber makes no mention of the influence of context which may necessitate placing the same noun in two different categories. In some cases, the problem of nouns fitting into several categories could only be resolved by consulting the co-text of the node word of the concordance line.

For example, in concordance line (11) below, the noun **solution** was categorised as *Abstract* while in line (12), the same noun was categorised as *Concrete,* based on my consultation of the surrounding context. In Line (11), the noun *solution* is used in the *Abstract* sense since it refers to the following noun, *problems*. Similarly, in Line (12), the second head noun, *crystallization,* makes it clear that the first head noun, *solution,* is used in the Concrete sense to refer to a chemical solution. This is further justification for extending the pattern, *the N of (the) N,* to include the second head noun to investigate the first head noun fully.

(11) …located in areas that are difficult to reach and the ***solution*** *of these problems by doing some damage along the wall or ground.*

*(12)… in the wall as well as the husks by repeating the **solution** of salt crystallization and swelling and shrinking of…*

Similarly, in concordance lines (13) and (14), the noun **foundation** was classified as *Concrete* because of its co-text which provides the context in which it is used; it refers to the foundation of a building as is apparent from the use of co-text such as 'exist under the ground' and 'structure'. 'Foundation' in (13) and (14) is therefore not used in the Abstract sense (e.g. ***foundation** stage of this project*).

*(13)…without drilling because drilling is linked to the existence of the **foundation**, which always exist under the ground and it is…*

*(14)…and the monitoring of site conditions, and the structure of the **foundation** and ground work. One of the problems facing the…*

The second challenge of adopting Biber's (2006) framework is that some of the noun categories do not appear to be mutually exclusive. Nouns such as *calculation* and *knowledge* may be classified as *cognitive* or under the *Abstract* category. Nevertheless, the *Cognitive* category was retained in the noun categorisation framework adopted since this allows for a finer-grained analysis of the nouns, giving insight into the types of abstract nouns preferred by the two genres. As can be seen in the examples in Table 3.7, the nouns placed in the *Cognitive* category involve some kind of mental processing. This was the criterion I used to categorise the nouns as *Cognitive*; for example, in concordance lines (15) and (16), the nouns *determination* and *views* both relate to some kind of mental activity and were therefore classified as *Cognitive*. In (15), the contextual cues indicate that *determination* involves the calculation of the 'compressive strength of mortars' while in (16), *views* refer to the opinions of 'people with the skill and experience'.

*(15) This test method covers the **determination** of the compressive strength of cement mortars, using 50mm cube specimens.*

*(16)…is widely thought to be developed must take the **views** of the people with the skill and experience to meet the…*

Another category that presented some challenges in categorisation is the *Technical/Concrete* category, which is defined as 'tangible objects that are not normally perceived and/or cannot normally be touched' (Biber 2006: 198). Words such as *solution* and *sample* are placed in this category in Biber's T2K-SWAL list, but *sample* might also fit in the *concrete* noun category while *solution* could fit into either the *concrete* or *abstract* categories, depending on its co-text. I nevertheless decided to retain this noun category in the framework used for my analysis since there were some nouns in my sub-corpora such as *bill* in 'bill of contracts', *picture*, and *infection,* which did not fit into either the *Concrete* or *Abstract* categories on consultation of their co-text, and seemed to be more appropriately classified as *Technical/Concrete*. If we imagine a continuum between the *Concrete* and *Abstract* noun categories, these nouns would be placed somewhere in the middle of this cline. Examples of such nouns in my sub-corpora include *line, axis, schedule, data* and *arc*. The bolded nouns in concordance lines (17) and (18) might be classed as *Abstract* but they seem to be more appropriately classified as *Technical/Concrete* since they can be placed towards the *Concrete* end of the Abstract-Concrete noun continuum. The noun *line* is not in the T2K-SWAL list, *data* and *schedule* have been classified as *Technical/Concrete* and *axis* is an *Abstract* noun in Biber (2006).

In Example 17, *data* refers to something can be 'transferred through email' while in (18), *line* refers to an imaginary line which is not entirely abstract, because it is easy to visualize and draw this 'line of sight' which guides the vision of a person looking through a telescope.

(17)…faster, as well as easier in transferring the ***data*** *of the entire project through email.*

(18)…are fully automatic and when properly set up the ***line*** *of sight through the telescope is horizontal…*

In some cases, I placed the same noun in two different categories depending on their contextual meaning. In concordance (19), the noun *lines* has been placed in the *Concrete* nouns category since it refers to the construction of a concrete object, a railway line. The

same noun was categorised as *Technical/Concrete* in concordance line (18) for reasons explained earlier.

(19)…is 2018. Riyadh metro involves the development of six ***lines*** *(blue, red, purple, and yellow, green, orange)*

Some nouns which have been placed in the *Technical/Concrete* category by Biber (2006) had to be placed in the *Concrete* category because of their meaning in context. Hence the noun *particles* has been placed in the *Concrete* category in Concordance (20) since it refers to soil particles which can be seen and touched.

(20)…the soil by sieving and to determine the mass of ***particles*** *in each range. 3. Materials: Sieve Shaker and…*

Another disadvantage of Biber's (2006) framework is the lack of sub-categories of nouns which may facilitate a more fine-grained analysis of the sub-types within the main categories of nouns. For example, *Quantity* nouns denoting time such as *century, week* and *age* are grouped together with those like *temperature, heat* and *energy* which can be measured using standardized measures. However, since a proliferation of categories would reduce reliability, I decided to use these eight categories, which are fairly tried and tested.

3.7.2.3. Data-driven categories: 'Gerund' and nominalisations

The query used to retrieve instances of *the N of (the) N* pattern also retrieved *–ing* forms occurring in this pattern; these are sometimes referred to as 'gerundial nouns' (Huddleston and Pullum 2002). They describe these words fulfilling the function of nouns by taking pre-modifiers and post-modifiers and by being the subject and object of clauses. Since there were many instances of these types of 'nouns', and, as shown in Examples 21 and 22 below, they typically refer to processes, I decided to include them as a separate category termed *Process nouns*. Examples (21) and (22) are instances of such uses of *Gerunds* in the CS and MR texts. The words *paving* and *building* are verbs in form but take on the functions of nouns in these sentences.

*(21) distance: when there is a body adjacent to the **paving** of bridge pillar or a wall or shoulder supported mile or…*

*(22) …of tunnels and 40km of bridges, and therefore the **building** of (50 terminals) and eight marshaling yards. In total, this can...*

The identification of *(verb)+ing* forms as nouns by the tagger follows Huddleston and Pullum's (2002) definition of 'gerundial nouns'. The tagger classified these *(verb)+ing* forms according to their function in the sentence. These 'Gerunds' were tagged as nouns by *Sketch Engine* using the tagger *English 3.3 for TreeTagger pipeline v2* when they occurred as the first head noun in the *N of (the) N* pattern. When the concordance lines were examined, I found that they were mostly tagged as verbs when they occurred in the second noun position in the pattern *the N of (the) N*. They were classified as verbs when they were immediately followed by a direct object and fulfilled the function of pre-modifiers. If they occurred at the end of a clause, they were categorised as nouns but if they functioned as pre-modifiers to the subsequent noun, this noun was identified as the second head noun.

In concordance lines (23) and (24), the words *collecting* and *extracting* are tagged as verbs while in concordance lines (25) and (26), *manufacturing* and *monitoring* are tagged as nouns. In fact, these concordance lines (in which the 'ing' forms are tagged as verbs) were retrieved as part of *the N of (the) N* pattern because of the nouns following these 'ing' forms.

(23) …the importance of collecting rainwater and reuse of wastewater.

(24)…they can protect the ground water and decrease the expense of extracting the water

I also examined the inclusion of Nominalisations under the *Abstract* and *Cognitive* noun categories. They were manually identified and their frequency of use compared across the CS and MR assignments. This was done because a preliminary reading of the nouns in the Case Study texts had revealed that many of the nouns in the *Abstract* and *Cognitive* categories were in fact Nominalisations. Investigating the use of this linguistic feature seemed to be a useful line of inquiry to uncover whether the types of linguistic devices used to achieve complexity in the two genres are realised through differences in noun type use

across the two genres. The use of Nominalisations has been reported in achieving a type of 'information density' by Gardner, Nesi and Biber (2018).

The procedure described above for categorizing the nouns in my sub-corpus using the noun category framework adapted from Biber (2006), Biber et al. (1999) and my own data, illustrates the need to consider contextual cues during the process irrespective of the way these nouns were categorized in other corpora and the way they had been automatically tagged.

3.7.2.4. Analysis of noun types used by the MR and CS assignments

This Section describes the statistical tests used to determine whether the differences in the semantic noun categories between the CS and MR assignments within *the N of (the) N* pattern are statistically significant. The number of instances of each semantic category represented by the first noun in *the N of (the) N* pattern in each text of the two sub-corpora were counted and their counts per 1000 words were normed. The norming was done because the word count of each text is different and it was necessary to standardize the frequency for comparison with the frequencies of occurrence of each noun category among the texts. This had to be done before calculating the mean occurrences of each category across each genre. It was then possible to apply the z-test to each category to find out if there is a statistically significant variation in these means across the two genres.

I then sought to find out the preferred noun type sequence/s (i.e. combinations of categories of first and second head nouns) within this pattern in each of the genres. In other words, I wanted to investigate whether specific patterns such as *Abstract Noun + Abstract Noun* or *Concrete Noun + Concrete Noun* in *the N of (the) N* pattern are preferred by either of the two sub-corpora in comparison with each other. I therefore identified the noun category sequences within *the N of (the) N* pattern in each concordance line and normed their occurrence per 1000 words in each text to standardize the frequency counts (to enable comparison and calculation of means) before applying the z-test. Although there were 25 Noun-Noun category patterns in the CS and MR sub-corpora, many of them were very rarely

represented (e.g. just one or two instances of a pattern); thus only those 10 patterns which were represented in both genres were selected for the z-test (zeros can distort the results).

The parametric z-test was applied because the sample sizes are 'large' by statistical standards (the CS sub-corpus has 31 texts and the MR sub-corpus has 63 texts). When large sample sizes are involved, the z-test is applied because 'the requirement for normal distributions is less important' (Blaikie 2003: 194). The t-test assuming unequal variance (variance is the distance of each data point from the common mean of the data set) could not be applied because the results are less reliable when the sample size is 30 and above. As with most parametric tests, the confidence interval was set at 95% meaning that the probability of differences in means occurring due to chance is 5% or less (Blaikie 2003).

3.8. Conclusion

This chapter has discussed the research design adopted for answering the research questions of this study. The Corpus Linguistics approach exploits the powerful computational abilities of corpus software to quantitatively analyse the corpus for the frequency and distribution patterns of linguistic items. Qualitative analysis in the form of genre analysis and examination of individual texts, as well as the use of qualitative data sources such as the views of specialist informants and course documentation, complements the quantitative analysis of the corpus. The importance of context and foregrounding of the emic perspective is reflected in the research design of this study. The ethnographic dimension of the project is thus highlighted with the researcher's unique positioning in the research site adding to the 'thickness' of the description.

Merriam and Tisdell (2016) suggest that an 'audit trail' in the form of a detailed description of how the data were collected and analysed builds confidence in readers about the consistency and dependability of a study. In this chapter, I have tried to build an explicit audit trail by giving details about the data collection and procedures followed for analyses in this chapter. This can allow other researchers to carry out similar studies to produce comparable results. I ensured that the strictest ethical guidelines were met which involves protecting subjects from harm, ensuring the confidentiality of their identities, the idea of

informed consent and lack of deception. Institutional ethical guidelines were met by this study by obtaining ethics approval for each stage of the study from the Ethics Committees of the degree awarding university as well as the research site.

This chapter contains a detailed explanation of the data collection procedure used for collecting student assignments in this study. Processing of the student texts and interactions with subject lecturers are also described along with the other data sources used in this study. The Corpus Linguistics techniques used to retrieve Lexical Bundles, nominalisations, noun strings and *the N of (the) N* pattern from the two sub-corpora are described in detail.

Chapter 4

Genre Analysis of Case Studies and Methodology Recounts in CE

4. Introduction

This chapter sets out to describe how the assignments in OCAW-CE were categorized into genres and the genre analysis process followed to demarcate the moves and steps in the two primary genres in the Omani Corpus of Academic Writing-Civil Engineering (OCAW-CE) following the Swalesian tradition (see Section 2.2). A description of these processes and the associated findings reported in this chapter thus addresses Research Question 1, which seeks to identify and describe the move structure of the two main genres in OCAW-CE.

As explained in Section 3.4, each assignment was examined and mapped against its primary social purpose based on Nesi and Gardner's (2012) identification of the five broad social purposes of BAWE assignments (see Section 2.3.4). The main rhetorical sections of each assignment were then identified. Discussions with subject lecturers and course documentation informed the two processes of identifying the primary purpose and discerning the staging of each assignment. These processes helped to categorise the assignments in OCAW-CE according to Nesi and Gardner's (2012) genre classification system.

As I expected, in some texts, the presence of clearly demarcated sections provided by the students made it easier to identify the genre. For example, texts with clear and highly conventionalized sections such as the *Introduction*, *Methodology*, *Results* and *Discussion* (IMRD), could be identified as a 'Methodology Recount'. However, some assignments contained sections that were components of other genres according to Nesi and Gardner's (2012) genre categories. These assignments were classified based on their primary purpose and this step heavily relied on discussions with subject lecturers (see Appendix 4), consultation of assignment briefs (see Appendices 3A and 3B) and examination of student assignments (see Sections 4.2 and 4.4). For example, in the *Civil Engineering Practice* Module, one of the assignment types contains an extended Literature Review section. I categorised this assignment as a Case Study because the primary purpose of this assignment

was to analyse a single exemplar (a building) as was evident from the assignment brief and the introductory section in which the student had also identified the aim of the assignment. In such cases, where there was some ambiguity about the primary purpose of the genre, the two other raters (See Section 3.4.2) were consulted and a decision was reached after discussion.

The prevalence of textual features such as tables, formulae and diagrams also provided clues that informed the genre-labelling process. For example, a large proportion of some assignments consisted of formulae and calculations; even before I examined the assignment briefs, these indicated that the assignment might belong to the Exercise genre.

The processes described above helped me identify six main genres in OCAW-CE: Case Studies, Methodology Recounts, Explanations, Exercises, Site Investigation Reports and Manuals (see Table 3.1). I found 63 assignments belonging to the Lab Report genre in my corpus, more than any other text type in this corpus. The next step was to identify the two main genres in the corpus to answer RQ 1.

I identified the Case Study and Lab Report (which is equivalent to a Methodology Recount in the BAWE classification system) as the two most important genres in CE at this institution (see Appendix 1). In this corpus word count and the number of texts were not treated as of the sole means of identifying the importance of genres, since some subject lecturers contributed more and some contributed fewer texts from their modules (see Table 3.1). Although the Case Study, Explanation, and Methodology Recount genres were all comparably distributed across the semesters, the subject teachers in both CE programmes said that the MR and CS genres were particularly important for their students. The subject lecturers also pointed out that students had to know how to write Lab Reports in order to attempt other genres such as Research Reports and Site Investigation Reports.

Sections 4.1 and 4.4 explain the move structure of these two genres to address the first Research Question of this study. This involved undertaking a fine-grained analysis of the assignments belonging to these two genre categories by identifying the moves and steps

in each section of the texts; few studies have undertaken this type of detailed move analysis of student genres from a single discipline.

The results of the move structure analysis of Case Studies and Methodology Recounts are presented in this chapter in preparation for the discussion of the results of analyses of phraseology and nominal features undertaken in the later chapters on these two genres. The moves in the CS and MR assignments were named after the naming conventions followed by Parkinson (2017). The initial letter of the main section was followed by a number; for example, when the moves of the results section had to be named, I named it Move R1, Move R2 and so on. Since some of the moves and steps are not present in every assignment, it was necessary to identify the obligatory and optional moves and steps (see Section 2.2) based on certain criteria. For this purpose, I adopted Parkinson's (2017) approach in her move structure analysis of Methodology Recounts in BAWE; I consider as 'obligatory' only those moves and steps which occur in at least 80% of the texts and as 'optional' those moves which are in 50-79% of the assignments.

4.1. Identification of Methodology Recounts

The MR assignments in OCAW-CE had the purpose of displaying students' knowledge of the conventions and experimental procedures of Civil Engineering to their readers, following the IMRD (Introduction, Methods, Results, Discussion) macro-structure. This accords with Nesi and Gardner's (2012) description of the social purpose and sections (IMRD) of MR assignments. Subject lecturers at the research site also confirmed that the primary pedagogical purpose of this genre is to assess students' ability to report experimental findings using disciplinary conventions. This primary function of Methodology Recounts is also reported by Parkinson (2017) and Kalaskas (2013).

Nesi and Gardner (2012: 142) identify the function of Methodology Recounts as an experiment that 'is written up primarily for the module tutor who appreciates why this is an appropriate topic to investigate and is more interested in what the student does (methodology), what she finds (results) and whether she can write appropriately using the IMRD framework.' This was the case with the MRs in OCAW-CE where a major proportion

of the assignment (42%) was devoted to the Methodology section, through which students display their familiarity with the experimental procedures in CE for the benefit of their examiners.

The similarities and differences of Methodology Recounts with other genres is worth a mention. Unlike the 'Research Report' genre, Methodology Recounts do not contain a substantial literature review component, long introductions and numerous references, even though experimental recounts may be part of this larger assignment (Nesi and Gardner 2012). A further distinction is that, although Methodology Recounts and experimental RAs have the same macrostructure (IMRD), they have different purposes and audiences; RAs contextualize the questions being asked and emphasize their contribution to the field of study while the main focus of Methodology Recounts is on following the methods prescribed by the lecturer with no attempt to place it in a broader context and to emphasize the contribution of the work (Parkinson 2017).

Another purpose of this genre, according to subject lecturers, is to prepare students for more extended assignments like the 'Research Report' which students are expected to write in the final year of their programme. This function of Methodology Recounts has also been reported by Gardner (2012). Yet another purpose is to anticipate workplace assignments which will require knowledge of such kind of experimental procedures. Lecturers indeed clarified that CE practitioners should be familiar with experiments such as evaluating the suitability of a particular kind of soil for construction purposes or the strength of a certain sample of cement.

As noted in Section 3.2, subject lecturers had sometimes provided more than the three samples per assignment for this particular genre, which explains the larger proportion of texts belonging to this genre in the corpus. The MR genre, which constitutes 32% of the assignment types in the Civil Engineering programme and 12% in the Quantity Surveying programme in OCAW-CE, was identified by the subject lecturers from both CE programmes as a commonly assigned pedagogical genre. (The two CE programmes share five common modules, see Appendix 1).

4.2. Move Structure of Methodology Recounts in OCAW-CE

Tables 4.1 through 4.5 provide the rhetorical sections of MRs in CE with their move structure analysis to illustrate how they realise the communicative functions of these sections. In most cases, the main sections in the MR assignments in OCAW-CE were indicated by genre-based headings: Introduction, Methods, Results, and Discussion. Only four of the assignments contained Abstracts; subject lecturers informed me that they did not expect students to include them in Methodology Recounts.

I found that the main function of the Introduction in these assignments (see Table 4.1) is to introduce the study by mentioning the aim and giving some information about the concepts or definitions of terms related to the experiment. The first move, labelled 'Introducing the MR', has four steps as shown in Table 4.1: 'Stating the importance of the topic', 'Referring to established knowledge', 'Providing diagram', and 'Stating aim'. This move involves introducing the experiment; stating the objective of the experiment is an obligatory step under this move (Move I1 Step 4; 100%). This step was identifiable because of the use of words and phrases such as 'aim', 'objective', 'purpose' and 'to phrases' such as 'to determine'. The presence of this step in all the assignments reflects its importance in this move. Move I1, Step 2 (73%) includes references to established knowledge among the members of the discourse community and are expressed in the present tense. Whenever there is a reference to a scientific/technical concept, the student explains the concept or the technique. For example, if the aim of the experiment is to determine the 'bulking' of sand, an explanation of the phenomenon of 'bulking' invariably follows (e.g. *Bulking of sand is the increase in the volume of sand due to the presence of moisture*). This type of explanation of technical concepts makes it clear that the student is also aware of the genre's pedagogical purpose and target audience. Move I1 Step 3 involves providing a diagram to explain the theory discussed in Move I1 Step 2 (22%). Sometimes students refer to sources and describe the procedure involved.

Move I2, labelled 'Stating the method', has two steps: 'Stating the method' and 'Justifying method'. Stating the method (Move I2 Step 1) that will be used in the experiment is found

in 57% of the assignments. In one of the texts (see Move I2 Step 1), we see that the student has used 'implicit justification' (Lim 2019) for the procedure adopted (*The theoretical discharge is calculated using the standard equation derived from **Bernoulli's theorem**)*. That the shared background knowledge of the reader is assumed by the student is apparent in the way 'Bernoulli's theorem' is mentioned but 'not everything is spelled out' (Giltrow 1994: 130). It is clear that explicit justification about the merits of the procedure being adopted is not necessary because both the reader and writer are disciplinary insiders with shared knowledge of what this concept (*Bernoulli's theorem*) implies. This type of justification of the procedure or method being adopted is a rare move and only one of the texts employs this type of 'implicit' justification, indicating that pedagogical genres may require explicit justification because their basic communicative purpose is to demonstrate knowledge. Two of the texts (3%) contain justifications for the methodology being used and provided diagrams to explain the concept or methodology.

Moves and Steps	**Examples**	**Percentage of assignments**	**No. of assignments**
Introduction			
Move 1 (Move I1) Introducing the MR			
Step 1	**Stating the importance of the topic** • *It is useful for locating drains, rods, dams and other earth structures.* • *The shear strength is one of the most important engineering properties of a soil, because it is required whenever a structure is dependent on the soil's shearing resistance.*	33.3	21
Step 2	**Referring to established knowledge** • *The phenomenon of capillary effect can be explained microscopically by considering cohesive forces and adhesive forces.*	73	46

	• *Finding the density of a metal sample can help to determine its purity.*		
Step 3	**Providing diagram**	22	14
Step 4	**Stating aim** • *The **objective** of this lab experiments is to find out the tensile strength of the given steel specimen.* • *The **purpose** of this experiment was to determine the dynamic viscosity of a given fluid*	100	63
Move 2 (Move I2) Introducing the Method			
Step 1	**Stating the method** • *Testing proceeds by displacing the lower half of the split box and measuring the horizontal shear forced (T) transmit through the soil to the upper portion of the box.* • *The theoretical discharge is calculated using the standard equation derived from Bernoulli's theorem.*	57	36
Step 2	**Justifying method** • *Soil particle size distributions test is faster than other test to determine the engineering properties of permeability, compressibility and shear strength*	3	2

Table 4. 1. Move Structure of the Introduction Sections of MR assignments in OCAW-CE

The steps in the methods section include listing the equipment or samples used in the experiment (M1 Step 1), and detailing the procedure (M1 Step 2) (see Table 4.2). These two main steps are also reported in studies that examined the methods sections in RAs in other disciplines as in Stoller and Robinson (2013) who described the moves and steps in Chemistry RAs. The function of the methods stage seems to be primarily to display knowledge since the reader is the lecturer. As indicated in Table 4.2, in almost all of the assignments, the details of equipment and samples are provided in the form of a list. The procedure (Move M1 Step 2) in most cases is detailed using imperatives and modals of obligation with very few texts using the past tense. The use of the passive voice is noticeable in some of the recounts. Students mainly use three linguistic devices to recount experimental

procedure: (1) imperatives (e.g. ***Take*** *an empty beaker*); (2) modals of obligation 'should' and 'must' (e.g. *The tool* ***must*** *stay vertical to the surface of the cup as groove is being made*); and (3) predicative adjectives controlling *to clauses* (e.g. *…for the removal of voids* ***it is necessary to*** *have a concrete with good consistency and flow*). Hyland (2002: 225) describes these linguistic devices as realising 'directives' which may be used to instruct the readers to perform an action (as here), refer to another section of the text or understand an argument from the writer's point of view. He examined a student corpus consisting of 64 research reports and found that more than half of all directives were employed by students to 'demonstrate the writer's mastery of practical research techniques'. These findings are consistent with mine regarding the use of these linguistic devices (imperatives, obligation modals, predicative adjectives controlling *to clauses*) by OCAW-CE writers in Methodology Recounts, illustrating the primary pedagogical purpose of this genre. Directive use in OCAW Methodology Recounts do not have the force of imposition they would have had if they been used for other rhetorical purposes such as directing the reader to follow a particular line of argument. Here, they mainly describe procedures, which, according to Hyland's (2002: 225) theory about the force of imposition, has 'no potentially face-threatening implications' since there is a clear understanding of the primary purpose of their use.

When subject teachers were asked about the use of imperatives in Methodology Recounts, they said that they did not give explicit instructions to students about how to write up their procedures. Many of the subject lecturers, in the course of my frequent encounters with them, have informed me that they disliked the use of first and second person in the 'procedure' section and had told students to avoid their use. Although the staff of the Centre for Academic Writing teach students the use of imperatives and the passive voice to recount procedures, it appears that most students prefer to use imperatives. This preference may possibly be attributed to the relatively easier form of imperatives rather than passive forms for second language speakers. The subject lecturers stated that as long as the content was appropriate and the first or second person was not used, they did not pay attention to the type of linguistic device used by the students to realise the communicative function of detailing experimental procedures.

In the same step (M1 Step 2), adverbs of manner (e.g. *thoroughly, properly, carefully, immediately, precisely*) are used to tell the reader how the procedure needs to be performed. It is noteworthy that a very limited range of adverbs (five) are used but they seem to serve the purpose because these assignments are still able to meet subject lecturers' expectations. Apparently, the emphasis on the accuracy of the content supersedes any linguistic consideration in terms of grades awarded by subject lecturers at the research site.

The language devices preferred by the Omani students to recount experiments might not always be used by students elsewhere. Parkinson (2017), who examined Methodology Recounts written by British university students, reports that imperatives are not used by the majority of the students in the BAWE corpus; they mainly used past tense declarative clauses to recount experimental procedure.

Differences in the language choices of Omani as compared to British university students may be due to educational background and departmental conventions, and could reflect preferences in some countries in the 'expanding circle'. Generally, accuracy rather than elegance of expression is the priority of students in expanding circle countries where English is used as a lingua franca (see 2.6). Both subject lecturers and proficient students shared the understanding that as long as information is conveyed clearly, using a limited set of syntactic and lexical features would not be an obstacle to success.

A minority of students recounted the procedure using the first person; subject teachers, however, did not think that a mention of the agent performing the experiments was appropriate whilst recounting procedures. This finding regarding departmental preferences reinforces my view that assignments collected from a single institution and examined in-depth offer insights to EAP teachers and researchers about what subject lecturers value and students need to know.

A few students included diagrams to explain the procedure (M1 Step2). Some texts also contain background information about the procedure, usually in the present tense because it is part of established knowledge.

Moves and Steps	Examples	Percentage of assignments	No. of texts
Methods			
Move 1 (Move M1)			
Step 1	**Listing apparatus/equipment** • *Different type of sieves, balance: to measure the soil sample weight, sieve shaker, cleaning brush and mixer or blender.* *a) Measuring beaker* *b) Eureka can* *c) Density bottle*	93	59
Step 2	**Recounting experimental procedure** • *Wipe every example to a surface-dry condition, and uproot any free sand grains or incrustations from the faces that will be in contact with the bearing pieces of the testing machine.* • *Write down the weight of dry soil in balance (about 500 grams).*	93	59

Table 4. 2. Move structure of Methods Sections in MR assignments in OCAW-CE

As shown in Table 4.3, the results sections were identified by their primary function of announcing the results through various steps: 'Using location statements', 'Displaying results in a figure/table', 'Detailing data analysis process', 'Calculating results', and 'Reporting results'. Location statements (Move R1 Step 1) are used by students (49%) to point to the results obtained from the experiment. These location statements are used in the passive voice by students. As Swales and Feak (2012) point out, location statements contain a location element followed by a brief summary of the information in a figure or graph (e.g. *Figure 3 shows the relationship between…*). They point out that this device is a form of metadiscourse and is an expected linguistic device in academic writing. The fact that nearly

half the Methodology Recounts in OCAW have this element points to their importance in this genre.

Almost all the assignments display the results in the form of a table or graph (Move R1 Step 2), making this an obligatory step. Nearly 90% of the assignments display their results in tables, graphs and figures; the subject lecturers told me that the nature of the assignments made it necessary for students to do this, which explains the frequency of Location Statements in assignments. At the research site, the Centre for Academic Writing also teaches students the use of Location Statements, which may have contributed to the frequency of their use.

The results are then reported (Move R1 Step 5) along with the formulae or equations involved in the calculations (Move R1 Step 4). The results are usually reported using symbols such as equals signs and numbers and are the results of mathematical calculations; subject lecturers informed me that they did not expect students to 'write paragraphs' to report their results, which explains the replacing of verbs with special characters, symbols or numbers (e.g. *total force F is:* $F = \pi d\sigma$). Some students also gave some information on how the data are processed (in about 30% of the texts). Invariably, this step was included in the Results stage of student assignments.

Moves and Steps	Examples	Percentage of assignments	No. of texts
Results			
Announcing Results (Move R1)			
Step 1	**Using location statements** • *The above graph shows the relation between the sieve size and the percentage of sample passing through each sieve size.* • *As seen from table 1, wooden blocks can take loads approximately 9 times*	49	31

	when the load is in a perpendicular direction to its fibres.		
Step 2	**Displaying results in a figure/table**	89	56
Step 3	**Detailing data analysis process** • *…calculate the compressive strength as follows: Compressive strength= P/A* *The coefficient of permeability is standardized at 20°C, and the permeability at any temperature T is related to K20.*	19	12
Step 4	**Calculating results** • *Bulking factor of sand= [Maximum increased volume - 1] = [465/325 – 1] = 0.430*	87	55
Step 5	**Reporting results** • *The results show that the viscosity for glycerin is 1 x 10-3, for automobile oil is 3.7 x 10-4, and for cooking oil is 1.1 x 10-4…* • *Result: Dry density=17.663 g/ml.*	73	46

Table 4. 3. Move structure of Results Sections in MR assignments in OCAW-CE

As indicated in Table 4.4, the next section was identified as Discussion because its main function is to attempt to align the results obtained with the theory. In many of the assignments, the Discussion and Results sections are conflated because after the results are reported, there is some commentary about them (e.g. *Maximum bulking of sand occurs when moisture content is 5%*). In this example, the student is discussing the results of the experiment obtained on a sample of sand. The Discussion section has two main moves: 'Contextualising discussion (Move D 1) and 'Interpreting the results' (Move D2). The main function of the Discussion stage thus seems to be to interpret results (Move D2 Step 2) based on the presentation of selected findings (Move D Step 1). The main tense used is the present

tense since the information being discussed is factual. It is useful to report the tense being used since many students approach the Writing Centre at the research site with questions about the tense they should use in the Methods, Results and Discussion sections of their MR assignments.

Subject teachers informed me that they expected students to comment on the results obtained but that many students failed to do so. This is clearly the case since only 46% of the assignments contain this step (Move D2 Step 2). Rather than labelling this step as 'optional', this finding only highlights the pedagogical implications. Students need to be made aware of this step in the move structure and therefore even though it should be categorised as 'optional' based on my findings from the corpus, the specialist informants have named it as an 'obligatory' step.

Moves and Steps	**Examples**	**Percentage of assignments**	**No. of texts**
Discussion			
Contextualising Discussion (Move D1)			
Step 1	• *The Point at which an increase in strain occurs without an increase in the stress is defined as the yield point.* • *The viscosity of relatively high viscosity fluids, such as oil, glycerin, and castor oil can be determined.*	38	24
Interpreting results (Move D2)			
Step 1	**Stating selected findings** • *From the overall results of analysis, the Bulk density of sand is 13.765 KN/ m3 and Dry density of soil13.538 KN/ m3.* • *For this experiment KT equal to K_20 = 0.03 cm/sec.*	67	42
Step 2	**Interpreting results**	46	29

	• *There is some variation in values due to surface tension of water.* • *…which mean that glycerin is more viscous than the two oil, because of this the ball took long time when it fall in the glycerin.*		
Step 3	**Accounting for unexpected outcomes** **Stating sources of error** • *Possible source of error: Slipping of the steel at the testing machine grips*	11	7

Table 4. 4. Move structure of Discussion sections in MR assignments in OCAW-CE

In some cases, Discussion and Conclusion sections are combined while in others the Conclusion is independent. As shown in table 4.5, where there is a conflated Conclusion and Discussion section, some assignments (67%) provided the summary of the experiment by presenting selected results (C1 Step 1) with some assignments (about 77%) containing an interpretation of these results (C1 Step 2). The present tense is used to draw conclusions and interpret results in order to make general conclusions. Interpreting results (Move D2 Step 2) seems to be recycled here in some assignments where explicit conclusions are not drawn. Subject lecturers and the assignment briefs (see Appendix 3) indicated that students are not told to explicitly state the conclusions that can be drawn from the experiment although subject lecturers did mention that students are expected to do this. They informed me that even though students came to them for feedback on their assignment drafts and they noticed this, they did not point this out to students since they thought it was unethical to do so because it would be like 'giving the answer' to students.

Moves and Steps	Examples	Percentage of assignments	No. of texts
Conclusion			
Summarising the study (Move C1)			
Step 1	**Stating selected findings** • *…the total mass sample after sieving is 497g is less than total mass 500g before sieving where 3 g of soil is missing.*	67	42
Step 2	**Drawing Conclusions** • *The results of viscosity and specific gravity are within the acceptable values.* • *The Soil classification as per the British Standard (MIT soil classification) is Sandy Gravel (G S).*	78	49

Table 4. 5. Move structure of Conclusion sections in MR assignments in OCAW-CE

As can be seen from Tables 4.1 through 4.5, some of the steps are obligatory, some are optional and others occur rarely (the complete table including all the sections in the MR assignments is provided in Appendix 2A). This may be due to the differences in expectations among lecturers, which is reflected in the instructions given to students, or simply because lecturers assume that students know what is expected of them and therefore are not always explicit in their instructions. For instance, all the lecturers agreed that Step 2 or the 'Drawing Conclusions' move (Move C1 Step 2) is obligatory but some students did not include this step in their assignment. One of my specialist informants said that they considered justification of methods used as a method of evaluating students' knowledge of disciplinary conventions. Some others did not think this was required in Methodology Recounts. The move structure was similar across assignments but differences are apparent among the assignments in terms of the steps realized. One of the reasons for these differences can be attributed to the differing expectations and instructions provided by subject lecturers who

come from different educational contexts. Many subject lecturers have in fact pointed out that this 'inconsistency' in the instructions given to students by the lecturers needs to be resolved. Since the Move Structures of the MR and CS genres provided in this chapter have been influenced by the subject lecturers themselves and are based on authentic student assignments and course documentation, they can provide a reference point for structuring these type of genres on the understanding that the moves and steps are not prescriptive and therefore constantly evolve on the basis of changing needs and contexts.

4.3. Identification of Case Studies

The Case Study is an assessment tool used in all the semesters of study in CE except for Semester 3 (see Table 3.1). It is also the most widely assigned genre, with 34% (15 modules out of 43) of the modules in both CE programmes containing this assessment component (See Appendix 1). The literature on case study assignments is quite scarce except for a few studies (Nathan 2010, Nesi and Gardner 2012). Nesi and Gardner (2012: 41) describe the social purpose of the CS genre as demonstrating or developing 'an understanding of professional practice through the analysis of a single exemplar'. Nesi and Gardner also observe that they invariably involve real-life cases, because of which a large amount of contextual data needs to be considered by the student while engaging in case study assignments. They also discuss case studies in Business and Medicine in terms of broad moves involving the presentation of the case and the problem followed by recommendations. Nathan (2010), who discussed CS assignments in Business, draws the same conclusions as Nesi and Gardner regarding the functions and purposes of the CS genre.

My discussions with specialist informants at the research site also indicate that the CS genre is meant to prepare students for industry requirements. They deal with a single real world entity and contain recommendations. Regarding the audience being addressed by students, Nesi and Gardner (2012) report that British university students write as if they are practitioners, sometimes even giving themselves titles. However, the CS texts in OCAW-CE explicitly address the examiner with no attempt at mimicking a practitioner submitting a company report. This is shown by the proportion of text devoted to explanations and the

number of references provided (minimum of 10 items). Two of the texts also mention the learning outcomes of the coursework.

As noted in the introduction to this chapter, the moves and steps were identified in the CS assignments in OCAW-CE based on the functions and the language features used to realise them in stretches of text. Tables 4.6 through 4.9 provide the move and step structure of the Case Study assignments in OCAW-CE along with examples from the assignments (the complete move structure of CS texts is provided in Appendix 2B).

Unlike the MR assignments, the headings in CS texts are not widely accepted genre headings and therefore the sections tended not to be so immediately apparent. It was hence more problematic to analyse these texts since the heading provided rarely reflects the generic function of the relevant section. However, as with the MR move analysis process discussed in Section 4.1, discussions with subject lecturers, course documentation and close reading of the assignments helped me identify the sections and their functions, and the moves and steps used to realise them.

Most of these assignments (about 97%) usually contextualize the study first by presenting the background and the case being examined (Table 4.6 Move I1 Step 1). There is usually a problem or an evaluation to be done; therefore, recommendations are offered or evaluations are performed after the analysis of these problems or the aspects of the case. Subject lecturers clarify that the main purpose of assigning Case Studies is to get the student thinking about the case and analysing it in order to come up with practical solutions, in line with the sort of task that practising Civil Engineers are frequently called upon to perform.

The sections identified here are largely in line with Nathan (2010: 170) who identified three ‘core rhetorical moves’ or sections in Business case studies: Orientation, Analysis and Recommendations along with an accompanying Executive Summary or Abstract. The analysis of the move structure of the CS assignments in OCAW also shows that these three sections are obligatory elements in this genre in the CE discipline (100%). All the assignments contained an Introduction section followed by an extended discussion of the literature (or explanations without references), which can be equated with Nathan’s (2010)

'Orientation' section. The second section (Table 4.7) is the Case Description and Problem Statement, followed by a Findings section (Table 4.8) which sometimes contains interpretations, and a Conclusion section (Table 4.9) which has an obligatory recommendation step. Nathan's 'analysis' is usually found in the 'Findings' section in OCAW CS texts; the analysis, which subject lecturers viewed as the most important section of the assignment, involves explaining the details of the case and interpreting what they mean. For example, if students find that the technique of 'double glazing' is used to treat windows in the case study of a 'green building', they need to interpret this finding as being in accord with some of the criteria under the international benchmarks on green buildings. The recommendations are invariably stated in the Conclusion section although there may be a few 'advisory elements' or recommendations (e.g. *HCT Green-Nest is performing weak* [sic] *in this area as there are no treatment strategies for sewage.*…) in the preceding Analysis stage.

4.4. Move Structure of Case Studies in OCAW-CE

There are four main sections in the CS assignments: Introduction, Description of the Case Study, Findings, and Conclusion. About 39% of the assignments contained Abstracts. Although all the Abstracts stated the aim of the study and provided some background information, the majority of them did not report the methods used or mention the contribution of the study. About half of the abstracts report selected findings. The subject teachers said that they were not very satisfied with the abstracts written by the students. They also told me that students were not given specific instructions on how to write them but were asked to read examples of research articles. It seems that the implicit knowledge of subject teachers is not transferred to students, as a result of which students may have written Abstracts that do not meet subject teacher expectations. These findings have pedagogical implications of some import since it is apparent that students need to be made aware of how to structure abstracts and that subject lecturers need to be more explicit in their instructions to students about their expectation of the elements to be included in an Abstract in the CS genre.

As shown in Table 4.6, in the Introduction section, the first move (Move I1) involves 'Introducing the topic' by 'Providing background information' (Move I1 Step 1; 96.7%), 'Providing definitions' (Move I1 Step 2; 58%), 'Claiming importance' (Move I1 Step 3; 87%) and 'Stating the broad problem' related to the case study (Move I1 Step 4; 83.8%). Hence, all the steps in this move are obligatory. These are mostly in the present tense since they consist of general truths such as the presence of pollution or the lack of public transport in the Gulf countries.

Almost all the assignments state the aim of the assignment (Move I2) quite explicitly using the words 'aim', 'purpose' or 'objective', or implicitly using phases such as 'to find out'. Only 19% of the texts did not outline the structure of the assignment (Move I3). Most of the Introduction is in the present tense and mainly uses non-technical lexis due to the functions it fulfils; these language features helped me understand that these functions include providing a broad context in which the Case Study can later be situated. In some of the assignments, I found the move of outlining the structure in the next section of the text which I have identified as Description of the Case (CD) (see Table 4.7).

Moves and Steps	**Examples**	**Percentage of assignments**	**No. of texts**
Introduction			
Move I1 Introducing the topic			
Step 1	**Providing background information** • *Gulf Company Council (GCC) is a regional Intergovernmental political and Economic union consisting of all Arab states of the Persian Gulf, with the exception of Iraq.*	97	30

	• *The process of the transport planning will go through different phases:*		
Step 2	**Giving definitions** • *Sustainability is 'the open and unrestricted and conception of sustainable development' (Jacobs, 1999:33).* • *Corrosion means atmospheric oxidation of metals.*	58	18
Step 3	**Claiming importance** • *In order to cope with rapid urbanization, meet with the requisites of burgeoning trade, encourage sustainable development, and to diversify the region away from reliance on oil and gas in an attempt to boost industrial development…* • *Planning the transportation is very crucial in order to achieve sustainable transportation system.*	87	27
Step 4	**Stating the broad problem** • *Therefore, sustainable construction is considered as one of the main challenges of sustainable development (Hoffman, 2008).* • *This is leading to a high consumption of natural resources, increasing the emission of greenhouse gases and polluting the ecological system...*	84	26
Move I2 Stating the aim of the study			

Step 1	• *The main **objective** of this assignment is to write about mega project in Oman or in other country in the world.* • *This report **aims** to identify the principles of sustainability and the main factors that create sustainability in buildings.*	97	30
Move I3 Outline of the assignment			
	• *This paper has progressed in a sequence of phases: identifying the principles of sustainability through conducting a literature review; mapping out the important factors that create elements of sustainability in buildings in Oman according to the codes and standards; description of the selected case study which is chosen as sustainable project; analytical comparisons between the model criteria and the case study.*	80	25

Table 4. 6. Move structure of Introduction sections in CS assignments in OCAW-CE

The CD section (Table 4.7) was identified on the basis of its primary function, which is to describe the case after providing some context. The context consists mainly of a review of the literature and sometimes contains explanations and definitions with or without citing of sources (100%). This step (Table 4.7, CD1 Step1) is also mostly in the present tense except when referring to past events (e.g. a conference on sustainability). I condensed this explanatory stretch of text into one move (Move CD1) in my analysis since it has the function of providing background information about the topic while at the same time displaying the students' knowledge. It is effectively an expansion of Move I1 Step 1 with the only difference being that it is quite expansive and descriptive.

The next step (Table 4.7, Move CD1 Step 2) consists of describing the features or current situation of the case. These details made this step easily identifiable as they provided

locations and facts about the case in the present tense. CD 1 Step 3 includes specific details or numerical information (Move CD1 Step 3). About 85% of the texts used figures or graphs to explain the details of the case (Move CD1 Step 4).

The next Move (Move CD2) indicates the problem to be solved or the evaluation to be done; this move was found in all the texts. Problems are indicated by the use of words with negative meanings such as ‘accidents’ and ‘collision’ (see examples in Table 4.7 CD2 Step 1). The concept of evaluation is indicated by the mention of benchmarking the case study against international codes or standards (Table 4.7 CD2 Step 1).

Moves and Steps	**Examples**	**Percentage of assignments**	**No. of texts**
Introducing the case study			
Move CD1 Description of the case			
Step 1	**Explanations with/out citations** • *…some materials during their life cycle produces emissions that harms the environment and increases the temperature of the earth.* • *The weight of the soil used for reclamation causes the water in the clay beneath to move outward along the sand piles.*	100	31
Step 2	**Description of the case** • *The road is starting from petrol station and ending by the roundabout that leads to expressway.* • *The Metro Project comprises of six (6) Lines are first line starting from north of King Salman Bin Abdul Aziz Street and ending at Dar Al Badia neighborhood in the south and runs in the North-South direction along Olaya and Bath streets.*	100	31
Step 3	**Providing numerical data**	100	31

	• *The estimated cost is $14billion and the total cost (budget) was $20billion.* • *The R165S has a 50-hp Perkins diesel turbocharged…* • *In 2006, thirty seven million cubic meters were being dealt with and used again.*		
Step 4	**Displaying descriptions in tables/graphs/figures**	84	26
Move CD2 Indicating the problem/evaluation to be conducted			
Step 1	• *This road experienced many secondary levels of accidents like Rear-end collision and Sideswipe accidents.* • *According to the factors identified before, performance of this greenhouse is next analyzed to test the sustainability concepts on the building according to the international standers and codes that have been identified in tables 1 to 12.*	100	31

Table 4. 7. Move structure of Case Study description sections in CS assignments in OCAW-CE

As shown in Table 4.8, the next section is 'Results' or 'Findings' and contains only one move with three steps. Move R1 Step 1 reports the findings and most of the texts refer to benchmarks or the literature in with this step (100%). The findings are also displayed in a table, graph or figure in 88% of the texts (Move R1 Step 2). More than 85% of the students interpret the results obtained (Move R1 Step 3). It can be seen that all the steps in this section are obligatory.

Moves and Steps	**Examples**	**Percentage of assignments**	**No. of texts**
Results			

Move R1 Announcing the findings			
Step 1	**Reporting findings with/out reference to the literature/benchmarks** • *The table 3 shows the envelope of visibility and different design speed for full overtaking sight distances. From the results we have obtained that design speed for selected road is 85A.* • *…the building's primary source of energy is the Solar Panels.*	100	31
Step 2	**Displaying findings in tables/graphs/figures**	87	27
Step 3	**Interpreting the results obtained** • *Green-Nest is performing well in this criteria, as it has light fixtures, fittings and luminaires.* • *This satisfies many standards such as National Association of Home Builders (2014).*	84	26

Table 4. 8. Move structure of Findings sections in CS assignments in OCAW-CE

As indicated in Table 4.9, the Conclusion section contains the obligatory step of making recommendations and is found in 96.7% of assignments (Move C1 Step 1). This step was readily identified since it involves linguistic features such as the modals 'should', 'can' and 'must' and words such as 'solution' and 'suggestions' to make recommendations, as indicated by the examples found in Table 4.9. This step (C1 Step1) has also been found to be obligatory across disciplines by Nesi and Gardner (2012) and Nathan (2010). The subject lecturers at MEC pointed out that making recommendations based on the analysis of the case is key and that they gave higher scores to assignments having clearly articulated recommendations. Move C1 Step 2 involves restating the aim of the study and summarizing the study. This is mainly done in the past tense. Only one student assignment alludes to the limitations of the study, making this step an optional one.

Moves and Steps	Examples	Percentage of assignments	No. of texts
Conclusion			
Move C1 Concluding the study			
Step 1	**Making recommendations** • *Therefore, the construction company that has an innovative idea should send its laborers to take courses in other countries know about this area.* • *Therefore, there are many suggested solutions to improve the traffic process such as other types of intersection (single-lane dualling), traffic lights or roundabout to the busiest exits (SQU & Knowledge Oasis).*	97	30
Step 2	Restating the purpose/summarising the study • *This project discussed the innovation of construction in three buildings in Oman including…* • *In this report I explain more information about the Lupu bridge.*	100	31

Table 4. 9. Move structure of Conclusion sections in CS assignments in OCAW-CE

Regarding the format of CS assignments in OCAW-CE, there is variation among the texts within the sub-corpus, with all the sections of some assignments being written in continuous prose, while others have a listing and bulleted format, as found by Nathan (2010: 172). The CS assignments analysed here are in some ways different from the Business CS assignments examined by Nathan; he found no references while I found that all the CS texts in OCAW-CE had a reference list containing a minimum of 10 items. Differences in the disciplinary realisation of genres have been pointed out by Nesi and Gardner (2012: 29). For example, they found that CS genres in Business differed from those written by Medical students in

their staging and organisation, just as Lab Reports written by Physics students differed from those written by students belonging to Food Sciences. Differences between my findings and those of other researchers could be due to differences in the disciplines, or in the way different institutions realise these genres. As discussed in 3.1.1, the cultural and educational backgrounds of lecturers in CE at MEC may have resulted in different assignment expectations from those of lecturers at British universities.

Once again this contributes to my argument for focusing on the writing in a single institution for understanding of departmental requirements and the importance of comparison to similar disciplinary and 'expanding circle' contexts who may or may not insist on the inclusion of particular components in such genres such as a substantial literature review. The complete move structure of CS assignments in OCAW-CE is provided in Appendix 2B.

As can be seen from Tables 4.6 through 4.9, almost all the steps, except for Move I1 Step 2 are obligatory i.e. they occur in more than 80% of the texts. The main sections identified above (Introduction, Description of the Case, Results, Conclusion) are in accord with those reported by Nesi and Gardner (2012) and Nathan (2010), who both include setting the scene, introducing the case and the 'issue' and ending with recommendations or 'suggested solutions' as sections found in the CS assignments they examined. However, the identification of the steps forming each of these moves is a contribution of this study to the move structure analysis of student Case Studies in CE. It is not possible to compare these findings with other studies since no other study, to the best of my knowledge, has analysed the move structure of Case Studies in Civil Engineering.

4.5. Manuals

As noted in Section 3.2, one of the other less frequently occurring genres in OCAW-CE is the Manual genre, which fits Nesi and Gardner's (2012: 212) description of Empathy Writing as mimicking 'non-academic genres ostensibly intended for non-experts' fits the Manual genre in OCAW-CE. In the assignments categorised as 'Manual' in OCAW-CE, the students are asked to take on the role of a practitioner in the field and address a real audience instead of the examiner. The function of this genre is to prepare students for professional settings,

where, as practitioners, they may be required to engage in such writing practices. The subject lecturers also confirmed this. There are only three examples of the manual genre represented in OCAW-CE because this type of coursework is only assigned in one module. This genre is not represented in BAWE or in MICUSP, two of the largest systematically archived student corpora.

The Manual genre consists of three main sections: 'Introduction', 'The Organization', and 'Arrangements'. The Introduction section contextualises the work by claiming its importance (e.g. *Health and Safety is a fundamental consideration.…*) followed by a statement of the aim (e.g. *The aim of this policy is to provide a safe environment for everyone who is involved with our handling operations*) and emphasizing the company's commitment and sincerity to the general public (e.g. *We are firmly committed to generate a safe sustainable long-term environment*). The next section is labelled 'The Organization' and functions as an explanation, where key terms are defined and explained (e.g. *Accidents are known as the events that happens without a plan, those events causes harm to the health of people, and damage to environment or property*). The next section, labelled 'Arrangements' mainly recommends precautions to be taken and the procedure to be followed in the event of an unforeseen event. This section forms a major part of this assignment. The function of giving recommendations is usually accomplished by using modals such as 'should' and 'must'. The precautionary measures already taken by the hypothetical company are expressed using the passive voice (e.g. *All appropriate plans are taken by the company to avoid any outbreak of fire in workplace.…*) or by the use of imperatives (e.g. *identify, evaluate, decide*). The importance of these activities is indicated using words such as 'important' and 'ensure' and modals like 'should' and 'must'. Many of the instructions are in a list and bullet format.

4.6. Site Investigation Reports

Site Investigation Reports fulfil the broad social purpose of 'preparing for professional practice' (Nesi and Gardner 2012: 36). Subject lecturers informed me that students will be expected to write Site Investigation Reports once they become practicing engineers and that

this assignment was intended simulate this workplace genre. As can be seen in Table 3.1, there are only three samples of this genre represented in OCAW-CE since it is only assigned in one module. The site investigation report is another genre which is not represented in BAWE and MICUSP; this may be because the CE discipline is not very well represented in these corpora.

Although this type of assignment is labelled Site Investigation Report, the assignment does not instruct the student to role play the part of a practitioner but explicitly states that students are required to write a literature review and explain the importance of the activities in their 'technical report'. Students are expected to visit a 'site' from where they are required to gather samples for conducting laboratory experiments. They then submit the report, which, according to subject lecturers is known among their discourse community members as 'Site Investigation Report'. In one sense, this genre has the characteristics of both Lab Reports and Case Studies, since experiments are conducted and a real-life exemplar is investigated. This genre, however, deserves a separate category since its purpose and functions are different from Case Studies and Lab Reports. The purpose of this genre is to evaluate a site based on the literature as well as on the basis of lab experiments on samples collected from the site.

Regarding the main sections in this genre, the first section is an Executive summary stating the purpose of the study along with selected findings and conclusions drawn. The function of the following Introduction stage is to restate the aim and provide explanations of main concepts. The next stage is usually a short one-paragraph Methods stage which details the methods used to collect the samples from the site (e.g. *Middle East College took Civil Engineering students on a site visit to Al-Khoudh Valley to collect different rock samples and study the earth's deformation*). The largest section of the assignment is the 'Data Analysis' stage, which provides the findings. There is little discussion of the findings (e.g. *This rock is a sedimentary rock. It has thin defined layers*) although there is some explanation or contextualization (e.g. *It is composed of angular quarts, feldspar and rock fragments*), which clearly fulfils the pedagogical purpose of the genre – to show understanding of the subject. The short Conclusion stage (averaging about 150 words) includes a summary and also has

the function of presenting conclusions from the results reported previously (e.g. *The main findings were: the most common rocks in site were sedimentary rocks*).

4.7. Explanations

The assignments categorised as belonging to the Explanation genre in OCAW-CE focus 'less on activity and more on theories and concepts' (Nesi and Gardner 2012: 35). These assignments have the purpose of displaying students' knowledge of concepts, theories and techniques/procedures in their subject area. For example, students are asked to write about the different materials used for construction, the factors that may affect their durability or performance and the methods of constructing certain items (e.g. walls, fire resistant buildings). Therefore, they fit the broad social purpose of 'demonstrating knowledge and understanding' (Nesi and Gardner 2012: 36).

In OCAW-CE, the Explanation genre occurs only in one module in the Civil Engineering module while it is quite a widely assigned one in the Quantity Surveying and Construction Management Programme (in six semesters across nine modules). This points to differences in preferred genres even between two programmes of study in the same discipline.

The Introduction stage contains the main moves of introducing the topic, stating the aim of the study and providing the outline of the assignment. The aim of the assignment clearly points to the pedagogic purpose of the genre (e.g. *The objective of this assignment is to know about monitoring and controlling of the construction work*). There is one extended stage which provides the explanations with references to the literature. The Conclusion stage sums up the main points included in the explanation stage. Some students also explicitly point out the pedagogic purpose of the genre in their conclusion (e.g. *In this assignment, learned [sic] many things about controlling and monitoring.…*) and claim importance for the topic they have discussed in their assignment (e.g. *The monitoring and controlling is very important in any project.…*).

Regarding the language features of this genre, both descriptive paragraphs and bulleted lists are found in these assignments in OCAW-CE. For example, there are descriptions of materials (e.g. *It is the oldest of building materials, and although primarily a place of concrete and steel construction, is still used in a large number of buildings*) in continuous text but when types of materials are listed or procedures are to be followed, they occur in bulleted lists. In most of the assignments, tables are used to display information so that students can explain it. This is a feature of the explanation genre also reported by Nesi and Gardner (2012).

4.8. Exercises

The main purpose of the Exercise genre is to display knowledge of concepts and use disciplinary procedures to solve short problems. The texts mainly consist of short answers (which are in effect explanations) or solutions to questions using numerical data and calculations. Nesi and Gardner (2012) also report that Exercises in the BAWE corpus contain short numbered responses to questions with little descriptive language. However, these researchers report that this genre tests 'basic' knowledge and skills which points to this type of coursework being assigned in the early semesters of study. This does not seem to be the case in OCAW-CE; this genre is mainly assigned in Semesters 5, 6 and 7 (see Table 3.1). The subject lecturers explained that the highly technical calculations could only be handled by students in higher semesters and that they would need to know how to perform them before they started their final extended project.

Imperatives are used throughout the assignment as most of the information deals with procedure to be followed (e.g. *Fifth, insert the project resource by Clicking (View tab/Resource Sheet) and update the information*), or explain methods of calculation (e.g. *Assume % Loss of pre-stressing after transfer*). The items are defined using the present tense (e.g. *A shallow foundation is a type of foundation which transfers building loads to the earth very near the surface*). This genre does not seem to offer much scope for detailed move analysis since they are mostly short pieces of text with the bulk of it constituting numerical data in the form of equations and expressions followed by calculations.

4.9. Conclusion

This chapter summarizes the process of categorizing the assignments in OCAW-CE into genres based on Nesi and Gardner's (2012) genre families. The genre of each assignment was identified based on the identification of the primary social purpose and staging of each assignment, consultations with subject lecturers and course documentation. Six genres were thus identified in OCAW-CE: Case Study, Methodology Recount, Explanation, Exercise, Site Investigation Report and Manual. Table 3.1 provides the number of assignments categorised under each genre, the number of words in each genre category, the average number of words in each assignment and the distribution of genres across the semesters of study. The Case Study (CS), Methodology Recount (MR) and Explanation genres seem to be the most widely distributed across the semesters among the six main genres in OCAW-CE.

The two main genres identified in OCAW-CE are the Methodology Recount (MR) and the Case Study. The main purpose of the MR assignments is to check students' familiarity with disciplinary experimental procedures. The CS assignments, on the other hand, simulate workplace scenarios with the aim of familiarizing students with the type of writing they would need to do at the workplace. Detailed move structures of MR and CS assignments have been provided along with details of the percentage of texts that contain each Move and Step. This detail about the proportion of texts that contain a Move or Step can help us to decide whether a Move or Step is optional or obligatory. It can also have pedagogical implications if a particular Move or Step is considered obligatory by subject lecturers but are not found in the majority of the texts. These findings demonstrate the importance of making explicit to students the Moves and Steps of disciplinary genres. This necessarily involves close collaboration between subject lecturers and language specialists; the subject lecturers need to explain to the language specialists the purpose and scope of their assignments, so that the language specialists can identify the Moves and Steps and their associated lexico-grammatical features. This collaborative process will lead to the creation of discipline-specific learning material that is immediately relevant for students attempting various genres in their disciplinary areas.

The description of the move structure of the MR and CS assignments is a contribution of this study because this analysis is based on a single-discipline corpus containing a substantial number of texts belonging to each genre. This chapter has thus addressed Research Question 1, which concerns the description of the move structure of the two main genres in OCAW-CE.

This chapter has also reinforced my argument that Omani students can be successful by using a limited repertoire of linguistic resources. If students can be made aware of linguistic features associated with Moves and Steps, they can focus on these rather than a wider range of items. Interviews with the subject lecturers confirm that it is the accuracy of communication rather than the range of language features at the disposal of students for the communication of disciplinary concepts that decides the success of their assignments. The subject lecturers who are themselves L2 speakers of English have prioritized this in the writing of students and in these particular contexts in the 'expanding circle'. EAP lecturers and researchers should allow the subject lecturers to take the lead in explaining what they value in student writing so that academic writing provision can be tailored to these particular requirements.

Chapter 5

Lexical Bundles

5. Introduction

As discussed in Section 2.7, analyses of academic corpora have revealed the high frequency of lexical bundles (LBs) in written academic language (Ädel and Erman 2012, Biber et al. 1999, Biber and Barbieri 2007, Hyland 2012) and their contribution to disciplinary phraseology. Attempts have been made to investigate the developmental differences between L1 and L2 student writers (Adel & Erman, 2012, Chen & Baker, 2010), novice and expert writers (Cortes 2004), and L1 and L2 professional writers (Pan, Reppen and Biber 2016) in terms of LB use. However, there remains a dearth of studies that have investigated differences in LBs among genres in student writing in a specific Engineering discipline.

This chapter will attempt to answer Research Question 2 by discussing the results of the investigation of the LBs in the two major genres in OCAW-CE, Case Studies and Methodology Recounts. To address RQ 2a, it will first examine the frequency of LBs found in both these sub-corpora, followed by a structural and functional categorization of these LBs based on Biber et al.'s (1999) and Hyland's (2008) taxonomies (see Sections 2.7.3 to 2.7.5). To address RQ 2b, it will then compare these results to identify genre variation between the CS and MR genres. The design of the OCAW-CE, containing texts from only the Civil Engineering discipline, is ideal for this kind of analysis because it isolates the effects of genre and discipline.

5.1. Frequency of LBs in the Case Study and Methodology Recount sub-corpora

The frequency of LBs can be examined from various perspectives including the number of different types of LBs in the corpus and the overall frequency of LBs in the corpus (Conrad and Biber 2004); a comparison of the top LBs from each corpus can also be revealing. As explained in Section 3.6.1, the frequency threshold applied to the sub-corpora was adjusted to take account of their different sizes and the resulting effect of Zipf's law as shown by Bestgen (2019). A raw frequency threshold of 5 was applied to both sub-corpora to retrieve the LBs. The dispersion threshold was also set higher for the MR sub-corpus (5 as compared

to 3 for the CS sub-corpus) because of the larger number of texts it contains compared to the CS sub-corpus (the MR sub-corpus contains 63 texts while the CS contains 31 texts). These settings led to a comparable number of LB types being retrieved from the two sub-corpora: 111 LB types from the CS assignments and 126 LB types from the MR assignments.

If we then compare the normalized frequencies of the LBs between the MR and CS sub-corpora, however, we can see that the MR texts contain a far higher overall normalised frequency than the CS assignments, at 20,755 per million words (pmw) to 8,610 pmw, respectively (see Table 5.1). This seems to indicate that the MR texts are more formulaic than the CS ones, since the proportion of LBs is much higher. In a way, this is to be expected, since the MR genre can be seen as a more formulaic genre, consisting of descriptions of processes, objects used in experiments and descriptions of calculations. Hyland (2008: 42) says that formulaic language signals 'competent participation in a given community' and that 'the absence of such clusters might reveal the lack of fluency of a novice or newcomer to that community.' What he doesn't take into account, however, is the effect of genre on LB use, which seems to be shown fairly clearly here at least in terms of formulaicity. However, these frequency figures tell us nothing about the individual LBs used or, in fact, how they are used in texts, of course. This method does not reveal the appropriateness of LB use because LBs are examined out of their context. Questions such as these can only be answered through a complementary qualitative analysis, which is undertaken in Sections 5.2 and 5.3.

	Corpus size	**No. of texts**	**LB types**	**Raw Freq. of LBs**	**Normalised freq. (pmw)**
MR	39,748	63	111	825	20,755
CS	105,806	31	126	911	8,610

Table 5. 1. Proportion of LBs in the CS and MR sub-corpora

5.1.1. Comparison of LBs in CS and MR sub-corpora

It is interesting at this stage, in considering the influence of genre on the distribution of LBs and what it can tell us about phraseology, to compare the most frequently occurring LBs found in the MR and CS sub-corpora. These are shown in Table 5.2.

Top CS LBs	Hits	Top MR LBs	Hits
as well as the	41	*bulk density of sand*	21
is one of the	34	*the mass of the*	17
as shown in figure	24	*the weight of the*	17
the gulf cooperation council	22	*determine the mass of*	15
higher college of technology	20	*the number of drops*	15
the sultanate of oman	20	*dry density of soil*	14
the total length of	20	*the base of the*	13
in the construction industry	14	*this experiment is to*	13
gulf cooperation council countries	14	*in contact with the*	12
united arab emirates uae	14	*the volume of the*	11
kingdom of saudi arabia	13	*can in the oven*	10
the construction of the	13	*experiment is to determine*	10
the completion of the	13	*of this experiment is*	10
in the sultanate of	12	*is to determine the*	10
used in the construction	12	*aim to determine the*	10
at the end of	12	*building strength and stability*	10
as a part of	12	*into the porcelain dish*	9
the end of the	11	*the compressive strength of*	9
one of the most	11	*in the data sheet*	8
in addition to the	10	*moisture can in the*	8
on the other hand	10	*the moisture can in*	8
in the construction of	10	*in the oven for*	8
at the same time	10	*to a depth of*	8
for the most part	10	*the moisture content of*	8

Table 5. 2. Comparison of most frequently occurring LBs in the CS and MR sub-corpora (raw frequencies)

The most striking observation to be made about the LBs listed in Table 5.2 is that none of the LBs are shared across the two genres. This is highly unusual for comparisons of LBs across subcorpora of different types, particularly when we consider that these two genres are from the same discipline. Hyland (2008) compares across different disciplines representing hard, soft, applied and pure fields and finds that a number of LBs are shared by texts in all or at least three of the disciplines. Within Engineering, Rezoug and Vincent (2018) also find shared bundles in dissertations from different sub-disciplines. So this finding again shows

the value of taking a genre perspective to LBs. These findings should, however, be interpreted in the light of other factors that may have influenced the type of LBs generated. Different dispersion thresholds for the two genres would have yielded slightly different results, and moreover some students submitted up to six assignments, something which the thresholds did not take into account. Therefore the effect of idiosyncratic styles might have crept into the final list of LBs.

It is also interesting to note the meanings and structures of the top LBs in the two fields (this will be discussed in more detail in Section 5.3). As we can see from Table 5.2, the top three LBs in the MR sub-corpus, *bulk density of sand, the mass of the,* and *the weight of the* (which all exceed frequencies of 400 pmw) all have the *noun with of phrase* structure. They also share the function of performing calculations and describing objects, a part of the experimental procedures explained in the assignments, as shown by examples 27-29. Clearly, LBs which refer to quantities are important to MR writers.

(27) Then, I calculated to find ***the weight of the*** *samples and prepared the materials for*

(28)… then start to determine ***the mass of the sand*** *pouring cylinder + sand to the nearest gram and note it as (M1).*

(29) From the overall results of analysis, the ***bulk density of sand*** *is 13.765 KN/ m3 and Dry density of soil is 13.538 KN/ m3…*

The top LB in the CS sub-corpus, however, indicate a quite different pattern of use and have different structural qualities. The most frequent LB, *as well as the,* has a frequency of about 387 pmw, followed by *is one of the* (about 321 per million words) and *as shown in figure* (about 226 per million words). Two of these LBs, *as well as the* and *as shown in figure* are used as text organizers (see Examples 30 and 32), showing how students feel the need to use transitions to guide the reader through their CS texts, which are much longer than MR texts (3,285 words on average as opposed to 642 for MR texts). The frequent occurrence of visuals in Case Study assignments is apparent from the third most frequent bundle *as shown in*

figure. The third of these bundles, *is one of the*, is analysed by Hyland (2008) as having a descriptive function. In Example 31, however, it seems to be used to hedge the description.

(30) The traditional way Spread widely in all Omani villages in the form of artesian wells and springs and water ***as well as the*** *Aflaj irrigation system.*

(31) This tower ***is one of the*** *tallest towers in the world.*

(32) … also show same of different categories of risk ***as shown in figure*** *2: type of risk in construction project Figure 2: type of risk in construction project (Ahmed. n.d).*

This brief overview of the most frequent LBs in the two genres already shows how this approach can help us investigate the phraseology of these two student genres. I will now take this further by describing the structural properties of the LBs in the two genres.

5.2. Structure of LBs in OCAW-CE

As explained earlier, all the LBs retrieved from the CS and MR sub-corpora were categorized according to their structure using Biber et al.'s framework (1999) (see Section 3.6.2). During this process, two other researchers were consulted to make sure that there were no inaccurate classifications. This was a relatively straightforward procedure (Rezoug and Vincent 2018) although the percentage of bundles in the 'other' category (especially in the MR texts) indicates that some additional LB structures may need to be added to the framework of Biber et al. (1999) for less proficient or less experienced writers.

5.2.1. Structural categorization of LBs in CS and MR Assignments in OCAW-CE

Examples of the structural categorization of the LBs in the CS and MR sub-corpora in OCAW-CE are given in Table 5.3 with proportions for each sub-corpus.

Structure	Examples from OCAW-CE CS sub-corpus	%	Examples from OCAW-CE MR sub-corpus	%
noun phrase with of-phrase fragment	*the construction of the; the completion of the; the objective of this*	*22.9%*	*The main objective of; the top of the; two halves of the*	*36%*
noun phrase with other post modifier fragment	*construction and how to; tallest tower in the*	*22.1%*	*the soil into a; the moisture can containing; soil into a moisture*	*20.3%*
prepositional phrase with embedded of-phrase fragment	*as a part of; in the construction of; in the form of*	*14.9%*	*on the end of; at the bottom of; at a rate of*	*8.5%*
other prepositional phrase fragment	*at the same time; on the other hand; for the most part*	*16.6%*	*in the oven for; in the data sheet; of the object in*	*9.6%*
anticipatory it + VP/adjective phrase	*it is important to; it is expected that*	*1.2%*	-	*0%*
passive verb + PP fragment	*as shown in figure; used in the construction; can be used to*	*7%*	*is performed to determine; used to measure the; used as a gauge*	*2.5%*

verb + to	-	*0%*	*aim to determine the; performed to determine the; is to determine the*	*3.3%*
verb + NP	-	*0%*	*determine the mass of;* *and record the mass*	*3.3%*
Be + Adjective	*is available in the*	*0.5%*	-	*0%*
Be + NP	*is one of the*	*3.7%*	*is one of the*	*0.8%*
NP/verb+ that clause	*the off chance that; in the event that*	*1.4%*	*the cup that comes; water content so that*	*2.4%*
Other expressions	*as well as the; as one of the*	*10.5%*	*as follows compressive strength; the total maximum load*	*13.2%*

Table 5. 3. Structural categories of the LBs in the CS and MR sub-corpora of OCAW-CE with proportions in each corpus

In both the CS and MR sub-corpora the most frequent LBs are those containing noun phrases (NP) and preposition phrases (PP). These findings align with those reported by Biber et al. (1999) that NP and PP bundles are preferred in academic prose over verb phrase (VP) bundles, which are more frequently found in conversation. This being said, genre differences may account for the slightly higher use of VP bundles in the MR sub-corpus, thus pointing to genre as well as register as a factor that influences the choice of VP bundles.

As can be seen in Table 5.3, the four most frequent patterns in the CS texts are *noun phrase with of-phrase fragment, noun phrase with other post modifier fragment, prepositional phrase with embedded of-phrase fragment* and *other prepositional phrase fragment*. In the MR sub-corpus, the four most frequent patterns are the same as those found in the CS texts. This is interesting considering the finding above (see Table 5.2) that there is no overlap at all in terms of the most frequent LB forms; while structures show very close overlap, the forms used to realise them vary quite considerably.

The other patterns (*passive verb + PP fragment, anticipatory it, (be+) adjective, (be+) NP and (NP+) that clause, (Verb+) NP and (Verb+) to*) amount to quite a negligible percentage of the total LBs retrieved from the CS and MR texts compared to the top four patterns used by the students. About 10.5% of the MR LBs and 13% of the CS LBs do not find a place in the classification scheme, suggesting that this taxonomy may need to be adapted for accommodating student genres such as CS and MR. A few of the CS bundles retrieved, such as *the principles of sustainability* and *in the field of*, may have been borrowed from CS assignment briefs (see Appendix 3B) However, this overlap was not found with MR bundles, as MR assignment briefs only mention the topic of the experiment such as *sieve analysis* or *bulk density* (see Appendix 3A). Although some of the MR bundles contain parts of the name of the experiment, such as *the top sieve and*, and *the bulk density of*, these bundles were not found in their entirety in the MR assignment briefs.

5.2.2. Less frequently represented structures in the LB lists of CS and MR

It can be interesting to consider what is less represented as well as what is frequently found in a corpus. An interesting example of this is the structure *anticipatory it + VP/adjective phrase*, which is not found in the MR sub-corpus and is only represented by two LBs in the CS sub-corpus. This suggests that these students do not feel the need to express their stance using this structure, since these are engineering assignments which belong to the 'hard' sciences where personal judgement may not be expected. Another reason for could be that these students are unfamiliar with this linguistic feature, which tends to find greater use in the texts of more proficient writers (Chen and Baker 2010; see Section 2.7.10).

Another structure which seems somewhat under-represented is *Passive + PP* in both sub-corpora, although the CS sub-corpus, with 7% of LBs, contains nearly triple the proportion the MR sub-corpus. This finding is quite surprising considering that passive structures are reported to be preferred (Parkinson 2017) by British university students to recount the methods sections in Methodology Recounts. This structure has very low use compared to findings reported for ACE (Rezoug and Vincent 2018) and Hyland's EE sub-corpus (Hyland 2008). It is in fact the most frequent structure in Hyland's corpus (at nearly 30% of LBs), which contains texts written by Masters, PhD and expert writers, while it accounts for between 12-16% of LBs in Algerian Master's dissertations (Rezoug and Vincent 2018). The explanation for this finding can be sought in the observation that the use of this structure correlates with writer proficiency and experience (Hyland 2008); undergraduate students, as the least proficient group of writers would be expected to use it the least. However, it is important to remember that this does not take into account the distinct genre traits of MR and CS texts in an Engineering discipline which may lead to the scarcity of these type of bundles in OCAW-CE. As noted in Section 4.2, the use of directives (mainly through imperative verbs) has been adopted by the Omani writers to recount experiments; this sort of use can be seen from LBs such as *weigh and record the.*

It is noteworthy that both *anticipatory it* and passive structures help the writer to avoid the explicit mention of the first person (although the 'anticipatory it' structure does express stance without foregrounding the writer). The low frequency of *Passive + PP* and *anticipatory it* bundles in OCAW-CE should be interpreted in the light of subject lecturers' comments when they were interviewed about their expectations of student writing. Their comments clearly show their dislike of students using the first person in their assignments. Some of their comments are given below. These discussions were part of the preparations undertaken by the Writing Centre to support students in writing up their Methodology Recounts. Omani writers may therefore benefit from explicit instruction in passive and *anticipatory it* structures although subject lecturers apparently are not concerned about the linguistic feature used by the students as long as they do not use the first person.

Discussion topic: When you look at a student assignment, what is it that you dislike the most?

Subject Teacher 1: Using first person pronouns 'I', 'We' even in spite of telling them… write as far as possible in active voice, mixed tenses and so on.

Discussion topic: Do you find differences in the written work of Level 0 and Level 1 students?

Subject Teacher 2: Yes, I have found a lot of differences between the Level 0 and Level 1 students. The same students when they are in level 0 struggle a lot to write their reports because they are new to the college and academics and …mostly in CE they are new and …mostly in writing they have difficulties in English grammar and how to write academically and …mostly use 'I took this', 'I brought this'. We guide them not to use 'I' or 'You' while writing.

Comparing the percentages of the different LB structures used in CS and MR texts, differences can be seen in the *(PP+) of phrase fragment, (NP+) of phrase fragment,* and *other Preposition Phrase (Other PP) fragment.*

5.2.3. Additional categories

Of the other grammatical patterns found in Biber et al. (1999) but not found in Hyland's corpus, the pattern *(NP +) (verb +) that-clause fragment* is only represented by four LBs and the patterns starting with verbs (*(verb/adjective+) to-clause fragment, (verb+)that clause fragment* do not occur at all in the Case Study sub-corpus. This shows how in some cases, the LB structures used in the MR texts are different from those in the CS texts, differences which can reflect the rhetorical purpose of the genres. To take the example of *(verb/adjective+) to-clause fragment (e.g. aim to determine the),* LBs with this structure are mostly used to recount experiments, an important rhetorical section in MR texts. A further interesting point here is that, although Gardner, Nesi and Biber (2018) report that the *(NP+) that* clause pattern as a feature that is used for a discursive style of writing, it is used in the MR texts as part of specifying the location of or referring to an object in experimental

procedure (e.g. *water content so that*). These types of bundles mainly occur in the Methods sections (see Table 4.1; Move M1; Step 2).

The list of LBs generated from the MR sub-corpus revealed that additional categories could be added to the structural framework of Biber et al. (1999) for analysing LBs. It should be borne in mind that both Hyland's corpus (2008) and Biber et al.'s (1999) Academic Prose corpus did not contain either the CS or the MR genres. Finding some additional structural categories in the MR genre is indicative of the specific demands of this genre in terms of academic writing. I found 21 instances (528.32 pmw) of LBs with verbs (e.g. *weigh and record the, indicated by the testing, comes in contact with*), which reflects the fact that the MR genre involves the recount of experimental procedure. The structure *(Verb+) PP* or simply *(Verb+) other modifier* may be considered for inclusion in the taxonomy. Another type of bundle which may be considered for inclusion in the structural taxonomy for analysing LBs is the *Noun Phrase* pattern, which usually includes the names of equipment or samples used in experiments. I found 36 instances (905.7 pmw) of such noun phrases in the list of LBs retrieved from the MR texts (e.g. *a smooth uniform paste, nonabrasive non brittle material*). Not a single instance of such noun phrase LBs were found in the CS texts. This type of noun string comprising a head noun preceded by multiple pre-modifiers help to densely pack information into a few words. The CS writers apparently do not use this feature frequently enough because the nature of their task does not require them to do this.

MR LBs with *(NP+) other modifier* structure constitute over 22% of the LBs in CS and around 20% of those in MR texts. The *NP+other* structure in the MR texts mainly follows the pattern of the noun phrase followed by a prepositional phrase (*NP + PP*). Of these, 37% are used to describe the location of objects (e.g. *the sand in the, the block on the*), indicating the importance of location in MR texts. This finding could perhaps be attributed to the frequent references to objects and things in the form of lab equipment or samples such as *cup, moisture can* and *block* in the MR assignments. About half of the LBs belonging to the *(NP+) other modifier* category in the CS texts are used to refer to proper names (e.g. *Kingdom of Saudi Arabia, the Gulf Cooperation Council, Lusail Light Rail Transit*). A fairly large proportion of the LBs (about 9%) in the CS sub-corpus are proper nouns and I have

categorized these under the *(NP+) other* structure because they do not exactly fit into any other category. These proper names allude to the location or the names of the cases being analysed in the assignments, an important function of LBs in the CS genre. The addition of the Noun Phrase structure to Biber et al.'s framework (1999) may be considered based on this.

In the Case Study assignments, the proportion of LBs containing *other prepositional phrase* fragments is nearly double that found in the MR sub-corpus. Examining the LBs in the CS assignments revealed that many of these LBs are signalling phrases like *on the other hand, at the same time, in addition to the,* and, *as a result of* and text organizers such as *of this project is,* and *of the project is.* The Case Study assignments in OCAW-CE are more descriptive and much longer than the MR assignments which might explain the higher frequency of LBs realizing the discourse functions of transition markers and structuring elements.

5.2.4. 'Noun phrase with of' fragment, (PP+) of phrase fragment, and other PP fragment.

It can be seen from Table 5.3 that the *(NP+) of phrase* bundles constitute 36% of the total bundles retrieved from the MR sub-corpus in terms of structure and the proportion of bundles with this structure retrieved from the CS sub-corpus is around 23%). This structure is also well represented in Rezoug and Vincent's (2018) ACE corpus and Hyland's (2008) corpus. This is likely because, as Hyland (2008) notes, this structure is associated with reference to quantity, quality, location, size, and existence which are commonly found in Engineering texts.

Concordance lines from the CS sub-corpus containing the *(NP +) of* structure are given in Table 5.3 to illustrate this point. In the examples below, the node forms refer to location (*the top of the*), quantity (*the weight of the, the total length of the*), and description (*the structure of the*). The *(NP+) of* structure LBs in the CS and MR texts also realise the textual discourse function (*the main objective of, the aim of this*). The main function of the *(NP+) of structure* in the MR texts seems to be to indicate quantity (e.g. *the temperature of*

the, the two halves of). More analysis of this structure is undertaken in Chapter 6 to compare the use of this type of bundle in the MR and CS texts. As can be seen in the examples below, the LBs with *(NP+) of* structure are followed by nouns (*bridge, building, track, soil, route, fins, construction, tower*) belonging to a range of different semantic categories. For example, *bridge*, *track* and *fins* are Concrete Nouns while *construction* is an Abstract Noun. It can also be seen that all the *(NP+) of* structures in the examples below are preceded by the definite article. For this reason, in Chapter 6, the investigation of the *(NP+) of* pattern has been extended to *the N of (the) N* pattern along with an examination of the semantic categories of the nouns within this structure. Since LBs only retrieve a limited number of word strings because of the thresholds of frequency and dispersion, I conducted a wider corpus search of bundles with the *(NP+) of* structure, the results of which are described in Chapter 6.

The *(PP+) of* phrase is found nearly twice as much in the CS assignments (e.g. *in the construction of; in the form of*) compared to the MR texts (*e.g. on the end of; at the bottom of; at a rate of*). The frame *in the Noun of,* seen in the CS LBs having the *(PP+) of* structure, was identified as a 'productive frame' by Chen and Baker (2010: 35) in the professional writing they examined in the FLOB-J corpus. As can be seen from the examples, their functions vary in the two sub-corpora because of their differing communicative purposes; in the CS texts, they perform the function of description while in the MR assignments, they provide details of location and experimental procedure.

The *other Prepositional Phrase (other PP)* bundle is found twice as much in the CS sub-corpus (about 17%) and mostly fulfils metadiscoursal and location functions (e.g. *at the same time; on the other hand; in the Sultanate of Oman*). Part of the purpose of CS texts is to persuade the reader of the validity of the recommendations, which may explain why there is more reader engagement than in MR texts. Conversely, the MR texts are more factual and mainly have the purpose of demonstrating knowledge and reporting the results of experiments. The *other PP* LBs in the MR texts (e.g. *into the porcelain dish; in the oven for; into the top sieve*) mainly perform the function of location. This function, realised by PP bundles starting with the preposition 'in', was also noted by Cortes (2002) in her study of freshman composition essays.

These findings indicate that although LBs with the same structure are found in both sub-corpora, the functions they perform vary according to the genre.

1	tourists can ascent the observation deck at	**the top of the**	bridge. It is recognized that gravity as
2	and the site was selected for solar panels at	**the top of the**	building to provide shadows and reduce
3	rails and sleepers. Ballast should support	**the weight of the**	track and also the considerable cyclic
4	of sand piles put into soft seabed clay	**The weight of the**	soil used for reclamation causes the water in
5	and continues to Salwa in Saudi Arabia.	**The total length of**	the main railway route is around 2,177km. In
6	tower is equivalent five of A380 aircraft and	**the total length of**	stainless steel bull nose fins is 293 times the
7	duration of construction is 7 years and	**the total cost of**	the construction of the mega project, which
8	structure of the building. The stability of	**the structure of the**	tower will be through gyroscopic automated

Table 5. 4. LBs with (NP+) of structures in CS texts fulfilling the discourse functions of location, quantity and quality

Studies such as Biber et al. (1999) report that 60% of the LBs in 'Academic Prose' contain noun and preposition phrases. Many linguists (Pan, Reppen and Biber 2016, Römer 2009) have also claimed that writing with noun and preposition phrases are preferred by experts and advanced writers based on comparisons of student genres with published ones. Although it may be true that academic writing in general prefers the use of NPs and PPs, differences in their constituent elements (e.g. type of nouns) and analysis of genres not included (e.g. CS and MR) need to be done to refine such claims and provide greater insight. The results of my analysis are consistent with the more nuanced interpretation of the use of linguistic devices and their clustering which Gardner, Nesi and Biber (2018) associate with the situational variables of level, genre and discipline in student writing.

5.3. Functions of LBs in OCAW-CE

The discourse functions of the LBs in the CS and MR sub-corpora were examined using the classification scheme proposed by Hyland (2008) (see 2.7.5 for more details regarding this taxonomy). Hyland's three main functional categories, Research Oriented (RO), Textual Oriented (TO) and Participant Oriented (PO), correspond to the three linguistic macro-functions of language identified by Halliday (Hyland 2012). Hyland (2008) then further specifies the sub-functions under each main functional category. Although, as noted in Section 3.6.2, applying the framework creates some difficulties, it allows us to get an overview of how the bundles are actually used in these texts and to discern further differences between CS and MR genres in terms of phraseology.

5.3.1. Functional categories of LBs in CS and MR assignments in OCAW-CE

Examples of the functional analysis of the LBs applied to the CS and MR genres are given in Table 5.6. As can be seen from the examples provided, although the main categories are relatively unambiguous, some of these LBs may have overlapping functions as pointed out by Rezoug and Vincent (2018) (see Section 3.6.2 for more details). Therefore, I had to make decisions regarding the order of preference for categorization when more than one category was possible.

Table 5.4 shows the proportion of the three main functional categories of LBs found in the CS and MR sub-corpora. In both sub-corpora, the RO bundles clearly constitute the major proportion of the total number of LBs although particularly so for MR at just under 84% of bundles. The tendency for less advanced writers to use more RO bundles compared to the other two categories has been reported by Hyland (2012) but the design of the corpus he used might not to be suitable for conclusive findings since genre differences rather than proficiency levels may account for this preference; the findings for OCAW-CE suggest some genre influence in their broad choice of functional bundle. The overall preference for RO bundles among the Omani writers of both these genres seems to reflect the pedagogical nature of the genres, which requires them to demonstrate familiarity with experimental procedure and discipline-specific concepts.

Another reason for the preponderance of RO bundles compared to the other categories in the OCAW-CE texts could be a disciplinary influence as has been noted in the literature (Hyland 2008, 2012, Rezoug and Vincent 2018). Other studies of engineering texts (Hyland 2008, Rezoug and Vincent 2017) have also found a higher frequency of RO bundles than the other two main function LBs.

In terms of TO bundles, the CS assignments have a much higher proportion than the MR texts. The comparatively greater length of CS assignments may account for this finding. The tendency for more frequent occurrence of these bundles in longer pieces of writing by more proficient writers (RAs and PhD theses) has been pointed out by Hyland (2012). These findings from OCAW-CE indicate once again the influence of genre in the choice of bundles

as the two genres within the same discipline show variation in the proportion of TO bundles used.

When we come to PO bundles, the MR assignments do not contain them at all, while students show some stance and engage with readers in the CS texts. This relatively infrequent incidence of PO bundles in engineering student genres has also been reported by Rezoug and Vincent (2018), and may again reflect the nature of the discipline or the task, which does not seem to require students to show stance or engagement.

	RO bundles	**TO bundles**	**PO bundles**
CS	70.0%	26.2%	3.8%
MR	83.9%	16.1%	0%

Table 5. 5. Proportion of LBs by main functional categories in CS and MR texts

Table 5.5 shows the distribution of the LBs in the CS and MR sub-corpora under the sub-functions of the three main functional categories (RO, TO and PO). Examples of LBs realizing each sub-function along with their proportion of occurrence in the CS and MR sub-corpora are also provided.

Research Oriented Bundles

	CS Texts		**MR Texts**	
Location	*the end of the;* *the completion of the*	*7.2%*	*the base of the;* *the center of the*	*10.1%*
Procedure	*the implementation of the;* *can be used to*	*8.9%*	*take the volume of;* *performed to determine the*	*26.5%*
Quantification	*is one of the;* *the weight of the*	*6.9%*	*the mass of the;* *the volume of the*	*16.2%*
Description	*the structure of the;* *the design of the*	*12.7%*	*the surface of the;* *a smooth uniform paste*	*9.7%*

Topic	*the Sultanate of Oman;* *in the construction industry*	*33.4%*	*the grooving tool is;* *limit and plastic limit*	*21.2%*
Text Oriented Bundles				
Transition signals	*as well as the;* *in addition to the*	*8.7%*	*as well as the*	*0.7%*
Resultative signals	*as a result of*	*0.6%*	-	*0%*
Structuring signals	*As shown in figure;* *The objective of this*	*10.9%*	*The main objective of;* *The aim of this* *this experiment is to*	*15.3%*
Framing signals	*for the most part* *On the off chance* *in accordance with the*	*6.5%*	-	*0%*
Participant Oriented Bundles				
Stance features	*ought to be given; it* *is important to*	*2.7%*	-	*0%*
Engagement features	*is expected to be;* *it is expected that*	*1.0%*	-	*0%*

Table 5. 6. Distribution of LBs by sub-function in the CS and MR sub-corpora

The following sections will discuss the findings by orientation, starting with Research-oriented bundles.

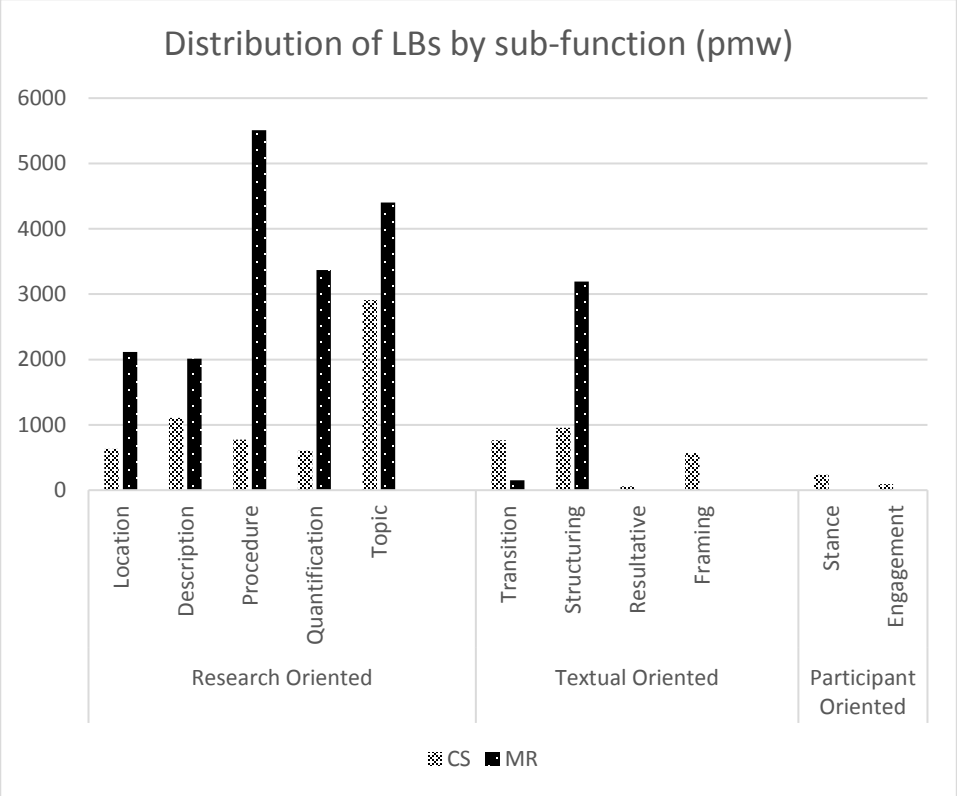

Figure 5. 1. Distribution of LBs by sub-function

5.3.1.1. Research Oriented Bundles

In terms of the RO function, it is clear from Figure 5.1 that by far the most frequent sub-function for CS texts is Topic, with Description a rather distant second. A different pattern is seen in the MR texts, where the sub-functions of Procedure, Topic, and Quantification are used the most frequently. As noted above, NP and PP bundles are used to realise RO functions in OCAW-CE; other studies of LBs in academic texts have also reported this association (Pan et al. 2016).

Among the RO bundles in the CS texts, Topic bundles are by far the most frequent, constituting about 33% of these bundles, followed by Description bundles at about 13%. Location (7%) and Quantification (7%) bundles are used less frequently and the Procedure

LBs constitute about 9% of the RO bundles. The high proportion of RO bundles in the CS sub-corpus can to an extent be accounted for by the nature of the tasks in CS assignments; they usually involve a considerable amount of description of engineering structures along with discipline-specific vocabulary used to analyse the case. The analyses of the case studies also involve some quantification, description of experimental procedure, and references to time of completion of a project, all of which are functions falling under the Research Oriented category. Another important sub-function of these bundles seems to be to describe objects, with LBs such as *the structure of the, the design of the, the surface of the,* and *in the form of*. Examples 34 to 37 illustrate these points.

(34) includes a piece of spiral stairs that reach to ***the top of the*** *tower. The completion of the first floor in the…*

(35) and then with compacted sand, and calculate ***the weight of the*** *loose sand (w3) and compacted sand (w4).*

(36) The transporting the goods by train after ***the implementation of the*** *Gulf railway network, it is expected that the…*

(37) wheel loader more than Gulfar Company. At ***the end of the*** *survey these differences depend on the type of…*

In the MR sub-corpus, Procedure bundles are the most frequent, constituting around 27% of RO bundles, followed by Topic (21%) and Quantification (16%) bundles. Description and Location bundles make up about 10% each of the total LBs. The reason for the relatively higher number of Procedure and Quantification bundles compared to Description bundles is because of the nature of the tasks attempted in MR assignments, involving the recount of experiments and associated calculations.

(38) The point on ***the cup that comes*** *in contact with the base should rise to a height of 10 mm.*

(39) Take four empty ***moisture cans with their*** *lids, and record the respective masses and can numbers on the data sheet.*

The *noun phrase with of-phrase fragment* pattern seems to be the most prevalent type of structure realizing the Description, Location, Quantification and Topic sub-functions of RO bundles (e.g. *the length of the, the weight of the, the total cost of the, the design of the,* and *the top of the).* This pattern is further explored in Chapter 6.

MR assignments contain far higher numbers of Procedure bundles compared to the CS assignments. This high frequency of Procedure bundles in MR texts is noticeable when compared to ACE (Rezoug and Vincent 2018) and Hyland's EE sub-corpus (Hyland 2008). This finding is hardly surprising as the 'methods' sections which constitute a major part of the texts in this sub-corpus (42%) consist of descriptions of experimental procedure. Nevertheless, it is noticeable that this reflects quite a high level of formulaicity in terms of how these methods are described.

With regard to Topic bundles, we can see from Figure 5.1 that the CS assignments contain higher overall numbers than the MR sub-corpus. Perhaps the fact that the MR sub-corpus contains short texts on many different topics may account for the frequency of these types of bundles. In the CS texts, an examination of the Topic bundles reveals that about 24% of them include place names like *in United Arab Emirates, Sultanate of Oman and* and *Kingdom of Saudi Arabia* and 29% include bundles containing the word *construction* preceded by the prepositions *of* or *in*; they are highly location-dependent. In the MR texts, on the other hand, the bundles usually contain the names of equipment, experimental samples or references to disciplinary concepts (e.g. *of the grooving tool, a smooth uniform paste, the compressive strength of*); these are more technical.

The frequency of Quantification bundles such as *the number of drops, the density of the, two halves of the, the volume of water, and the two halves of* in the MR sub-corpus point to the frequency of calculations and measurements in MR assignments. This, however, does not seem to be the case with CS assignments which have fewer calculations.

The frequency of Location bundles (e.g. *point on the cup, block on the end, at the base of, in the data sheet,* and *at right angles to)* in the MR sub-corpus is higher compared to the CS sub-corpus. These LBs are used to describe the location of objects in lab experiments. An examination of concordance lines containing the Location bundle *the end of the* from the CS sub-corpus of OCAW-CE reveals that this bundle is generally used to refer to a time period, although it is used once as part of the idiomatic expression *the end of the day.* As can be seen in Table 5.6, it is only once used in the sense of physical location (Line 9) as they are used in the MR texts. This finding regarding the variation in the use of Location LBs between the CS and MR texts can be attributed to genre variation.

1	nearly two million visitors paid at	**the end of the**	show at the end of October. In terms of the
2	the final outcome of the work at	**the end of the**	accounting period, it may be profit or loss.
3	(Jocelyn, 2013). By	**the end of the**	twentieth century, Middle East countries
4	et al 2007). Demolition stage: It is	**the end of the**	building life as this stage involves a lot of
5	and assets reasonably, at	**the end of the**	day, at a rate that does not trade off the
6	and the materials used in finishing	**the end of the**	project, it has the choice of all these
7	more than Gulfar Company. At	**the end of the**	survey these differences depend on the
8	to get information from them. In	**the end of the**	coursework it was concluded that only
9	that butt up to each other. If	**the ends of the**	interlock are not straight when they are
10	lintel is prescribed. Then again, at	**the end of the**	day repointing will suffice until the

Table 5. 7. Location bundles in the CS sub-corpus of OCAW-CE

5.3.1.2. Textual Oriented Bundles

Textual-oriented LBs comprise just over a quarter of all the LBs retrieved from the CS sub-corpus, but only around 16% in the MR sub-corpus. The higher proportion of TO bundles in CS texts could be explained by the length of these texts which necessitate the use of textual markers to help the reader navigate them. Hyland (2008) found that TO bundles occurred more frequently in the softer disciplines (Applied Linguistics, Business Studies). Perhaps the Case Study genre in Civil Engineering may be considered 'softer' than other genres because of the greater amount of descriptions and explanations necessitated due to the analysis of case studies in the light of their impact on the wider world.

The most frequent type of CS bundles in the textual-oriented category are structuring signals. These help to organize the text and signal to the reader about the stages of the text and their purpose. Some of the structuring signals in these texts are *the aim of this*, *the*

objective of this, and *this case study is*. The relatively high frequency of Structuring signals such as *as shown in figure* point to the numerous figures and tables that characterize these texts. Nesi and Gardner (2012) also found that the student texts belonging to the 'Apprenticeship genres' (which include Case Studies) contain many visual devices for presenting information.

Although MR texts are much shorter than the CS texts, students make use of more structuring bundles than any of the other TO bundles. This may be attributed to the highly conventionalized organization of these assignments. As explained in Section 4.4, these assignments follow the typical IMRD format with the Introduction section always stating the objective or purpose of the experiment. Clear instructions are also given in the assignment brief (see Appendix 3A) about the different sections to be included (III. OBJECTIVES AND THEORETICAL BACKGROUND IV. METHODS/PROCEDURES V. RESULTS & DISCUSSIONS VI. CONCLUSIONS). In fact, 13 of the 18 LB structuring types realise this function of reporting the aim of the experiment (e.g. *the aim of this, objective of this experiment*). The remaining bundles contain parts of conventionalized headings (e.g. *observations, calculations and results, discussion and conclusion the*) indicating the typical headings used in Methodology Recounts in Civil Engineering and perhaps other disciplines too. It is worth remembering here that in the move analysis of MRs found in the previous chapter, the step of including the aim of the experiment in the introduction was found in all (100%) the assignments (see Table 4.1).

Transition signals with the additive function are the second most frequent in the CS texts (e.g. *in addition to the, in addition to that, as well as the, as well as its).* All except one of these TO bundles (*on the other hand*) fulfil the additive function. There is only one instance of a Transition bundle in the entire MR sub-corpus and there are no resultative or framing signals. This may be because of the short texts that make up the MR sub-corpus with clearly defined headings. This type of organization may not possibly require the writer to guide the reader through the text using transitions.

One example of a resultative transition marker, *as a result of*, is seen in the CS sub-corpus. The fact that only one type of resultative signal, *as a result of* is used by students should be of pedagogic interest to the EAP teacher as students may lack the linguistic devices to report results using a variety of techniques. It is surprising that there are no resultative bundles found in the MR assignments in spite of the fact that there is a separate Results section in all these assignments. An excerpt from the Learning Outcomes specified in an MR assignment brief given below is evidence of the importance of results in Methodology Recounts.

> *Carry out laboratory investigations and draw conclusions from experimental data on materials and present clear reports on the results and conclusions.*

Result bundles are missing in MR assignments because students report their results by replacing words by calculations and figures as discussed earlier in the move analysis of MR assignments (see 4.2) and also because less than half of the MR assignments interpret the results. The move analysis of texts also supports subject lecturers' comments about inadequate interpretation of results of lab experiments. Sometimes, the interpretation of results is realised through word sequences that are not recurrent enough to be retrieved as bundles (e.g. …*which mean that glycerin is more viscous than the two oil, because of this the ball took long time when it fall in the glycerine*.). Similar factors account for the lack of Resultative bundles in the CS sub-corpus, which contains only one instance of this bundle.

The presence of Framing signals in the CS assignments might be due to the expansive style of writing employed for this genre. These texts are much longer than the MR texts and require students to specify conditions and make comparisons or assess the case against international benchmarks, conditions, or frameworks in the course of studying their case (e.g. *in accordance with the terms and conditions of the relevant contract e.g. JCT standard building contract, in accordance with the frameworks and systems for scientific development, in the event that the tender period is short*). These examples demonstrate how the framing signals *in accordance with the* and *in the event that* is used to refer to international benchmarks (*JCT standard Building Contract*) or conditions (short tender period).

It is evident that the CS assignments show a marked preference for TO bundles compared to the MR texts. The discussion above has suggested that genre and its associated task requirements influence the distribution of TO bundles.

5.3.1.3. Participant Oriented Bundles

Participant-oriented LBs, which Hyland (2008: 18) categorizes into the two sub-functions of 'stance' and 'engagement', are distinguished as follows:

> *…stance concerns the ways writers explicitly intrude into the discourse to convey epistemic and affective judgements, evaluations and degrees of commitment to what they say, engagement refers to the ways writers intervene to actively address readers as participants in the unfolding discourse.*

According to Hyland's definition, 'stance' is primarily used by writers to position themselves and thus convey their judgement or evaluation of a proposition while 'engagement' involves addressing and engaging in dialogue with the readers of the text.

PO bundles were not found in the MR texts in OCAW-CE while under 4% of the LBs in the CS texts realise the functions of stance and engagement.

There could be many reasons for the sparse use of PO bundles by the Omani students in OCAW-CE compared to Hyland's writers. The first reason is the inability of corpus methods to identify all instances of 'stance' and 'engagement'. Another reason for the absence of these features (if indeed they are absent) could be because the Omani students lack the linguistic devices to express opinions. Yet another reason could be that students may have the language skills to make recommendations, but may lack knowledge of the CS genre. Since stance and engagement features in the CS texts are usually required in the 'Recommendations' (Table 4.2; Move C1 Step1), the reason for the lack of this feature in the CS texts could be that students do not give recommendations in their case studies. Perhaps Omani students do not require training in how to express stance and engagement strategies more generally, but just some advice on how to recommend particular actions because they are not aware of the requirements of the genre. The subject teacher confirmed that students

need to come up with recommendations in the conclusion of their assignment but also implied that in addition to not being able to express themselves in the English language students lack critical thinking skills because of which they are unable to come up with recommendations.

The following comment was made by a subject teacher on Case Study type of assignments clearly implying that students should adopt some stance and engagement as they are required to make their 'own recommendations' which require 'critical thinking':

> ***critical thinking*** *is involved in this, so students should analyse the buildings and analyse the case studies and all …so they have to gather some information and they have to reproduce it on their* ***own recommendations***.

Subject lecturers apparently consider the expression of stance and engagement important in student assignments as can be seen in the following comment from a Civil Engineering lecturer on his expectations of student writing in Methodology Recounts (MR genre). Students are required to 'discuss' and draw conclusions by applying their 'own judgement'. The subject lecturer commented that students need to interpret the results by evaluating them. This implies that students are required to use these features of stance and engagement.

> *so in the Lab Reports students are…they are asked to have their own soil samples, conduct some soil tests in the lab, then* ***discuss*** *and they also need to provide conclusions based on their* ***own judgement*** *regarding the results.*

Another reason for the lack of PO bundles could be the emphasis on the communication of information over argument and stance-taking in the hard sciences (Nesi and Gardner 2012: 72, Neuman, Parry and Becher 2002: 408). Although some genres in the hard sciences do not require their use, there are certain genres even in the so-called 'hard sciences' in which argument and stance are prominent. Clearly, the CS and MR genres do require students to express their opinions and engage with their readers though their degree of importance may vary since the MR might tend towards the purpose of communication of information and drawing of conclusions based on undisputable hard facts. Speaking about the requirements

of the MR genre, the subject teacher did confirm that what he meant by 'own judgement' is the application of theoretical knowledge to the results of the experiment which the student is not expected to question.

Hyland (2008) reports that PO bundles are found largely in expert writing and are not preferred by student writers in his corpus. He points out that this could be due to a reluctance to express strong opinions on the part of the L1 Chinese students he investigates. These conclusions do not seem to take into account genre factors; the Research Article genre might require different linguistic devices than those expected in student genres. Hyland (2008) does not consider genre variation as one of the reasons for the students not attempting to demonstrate stance or engagement in their writing.

Comparing this lack of stance and engagement LBs in the OCAW-CE with ACE and Hyland's corpus, it can be seen that these other corpora use PO bundles much more frequently that the Omani writers in spite of their texts being in the Engineering discipline; however, it should be recalled that the genres represented in them do not contain CS or MR assignments. It is perhaps dangerous from this perspective to compare the type of genres represented in ACE and Hyland's corpus, being extended pieces of work and written by groups of writers who would be expected to exhibit stance and engagement. This comparison only serves to highlight the possibility of genre variation and even levels of study causing these differences. The fruitlessness of attempting to compare different genres and concluding that a certain group of writers need to write in a particular way is apparent; interpretations should be more sensitive to genre differences and consider the appropriateness of linguistic devices for specific genres. Studies that are not sensitive to these differences (Chen and Baker 2010, Hyland 2008) run the risk of distorting their conclusions because of the incompatibility of the sub-corpora they are comparing.

Biber and Zhang's (2018) definition of stance and engagement varies considerably from Hyland's; they consider stance as the expression of attitude or judgements of certainty using a limited set of lexico-grammatical features while 'engagement' also expresses epistemic assessments and attitudes but they can be integrated into long or short stretches of

discourse and sometimes also take into account factors like the expectation of the reader. Biber and Zhang highlight the limitations of using corpus methods to identify what they call 'engagement' devices because they claim that manual identification procedures would need to be used to accurately identify them. For example, stance markers may also require manual identification since discourse markers (e.g. *fortunately*), signalling nouns (e.g. *mistake*) and modals (e.g. *should*) and even 'particular words and phrases' (Biber and Zhang 2018) may be used to express stance. Any attempt therefore, to categorise PO bundles as Hyland (2008) and other studies have done (Durrant 2017, Rezoug and Vincent 2018), using this classification framework is restricted to identifying only those stance features occurring within LBs. This will certainly not be an exhaustive list of stance features used in texts by any means.

The results of the categorization of the MR and CS LB lists with regard to PO bundles should be considered keeping in mind these limitations of the corpus method in identifying stance and engagement features.

5.4. Conclusion

Through the analyses of the frequency, and the types of LBs by structure and discourse function in the CS and MR texts, which I conducted in this chapter, I have attempted to answer Research Question 2. The findings of my analyses have revealed that genre variation is an important factor influencing the frequency and types of LBs used. These students have obviously been successful as their texts have been selected for this corpus on the basis of teacher evaluation. We have seen evidence that they employ a relatively limited repertoire of phrases raising again the question of whether a limited but relevant repertoire of linguistic items might be the feasible EAP approach for students in 'expanding circle' contexts (Sections 2.6, 4.2).

In terms of structure, I found that although NP and PP bundles are preferred by both the genres, the MR genre makes use of more VP bundles, reflecting the experimental recounts found in all the texts. These findings offer some challenge to the claims by studies that have recommended that students should attempt to progress from a verbal or clausal

style of writing to a phrasal style (Biber, Gray & Poonpon 2011; Pan, Reppen & Biber 2016). These researchers claim that students should emulate the writing of expert writers without considering the fact that student genres have a pedagogical function and may not coincide with the purposes of expert genres. I have also pointed out that the structural taxonomy (Biber et al. 1999) which I used in my analysis was developed on the basis of corpora that did not include student genres like the CS and MR. I therefore recommend that additional categories like *Noun Phrase* and *Verb Phrase* or *(VP+) other modifier* should be included so that the taxonomy is more accommodating of pedagogic genres like CS and MR.

Analyses of the functional categories of LBs in the CS and MR assignments also indicate that genre is a major factor influencing the distribution of LBs. Hyland's (2008) taxonomy, which I adopted for the analysis of the LBs in the CS and MR assignments has enabled me to examine the LBs by sub-function and identify the categories that have contributed the most to the differences between the two genres. The main findings were that Procedure, Quantification and Location bundles were much more frequent in the MR texts compared to the CS texts and that the proportion of TO LBs higher in CS than in the MR assignments. I have also attempted to associate some functional bundles with the communicative functions of the steps under the broader moves of the MR and CS assignments. For example, PO bundles can be used to make recommendations in the Conclusion section of CS assignments (see Table 4.2; Move C1 Step 1) as they involve the display of stance. Thus RQ 2a on the structure and functions of LBs found in CS and MR assignments has been answered, and an analysis of these LBs has revealed their similarities and differences in the two sub-corpora, thus addressing RQ 2b.

As with the structural bundles, my findings suggest that claims in the literature regarding the variables of language proficiency, native-ness and discipline leading to differences in the use of LBs need to be re-examined from the perspective of genre.

The findings of this chapter and the conclusions drawn are based on comparisons between two compatible disciplinary sub-corpora in terms of genre. This corpus design isolates the effects of genre unlike other studies (Chen and Baker 2010, Hyland 2008) where

other variables (multiple disciplines, different genres, varying proficiency levels) interfere to affect the validity of their findings.

Chapter 6

Nominal Features

The CS and MR sub-corpora are phrasally rather than clausally complex and the normalised frequency of nouns in the two sub-corpora is more than twice that of verbs. This chapter, therefore, compares nominal features in the CS and MR assignments, and by investigating genre variation in this respect it will address RQs 3a and 3b.

Gardner, Nesi and Biber (2018) consider the situational variables of level, genre and discipline in their student writing corpus, and point out that there are two types of phrasal complexity in academic writing. The first type, described as 'compressed procedural information' (Gardner, Nesi and Biber 2018: 9) involves the use of pre-modifying nouns, common nouns, passives, action verbs, concrete nouns, and quantity nouns. The second type, described as 'informational density' (Gardner, Nesi and Biber 2018: 22) is characterized by nominalisations, attributive adjectives, abstract nouns and long words. These researchers report that the first type of density is usually found in science report genres such as Methodology Recounts and Design Specifications in the Physical Sciences. The second type is usually found in assignments written at the more advanced levels in the social sciences. They interpret the findings of their study to mean that writing in the hard sciences differs from that in the softer disciplines in terms of the linguistic features used to achieve phrasal complexity.

In this chapter I will focus on nominalisations, noun strings, attributive adjectives, and *the N of (the) N* pattern. The pattern *the N of (the) N* was selected for particular scrutiny because *of*-phrases are the most prolific prepositional phrases used as noun modifiers in academic prose, according to Biber and Gray (2016), and a review of the literature also reveals that this pattern is common across academic genres and disciplines; for example, Marco (2000) found that *the N of* was the most productive collocational frame in the medical RAs that he examined. In Chapter 5, I reported that *the N of (the) N* was the most frequent four word lexical bundle structure in both the CS texts (e.g. *the construction of the, the number of drops*) and the MR texts (e.g. *the mass of the, the bulk density of*). In this chapter

I used corpus query language rather than lexical bundle extraction to search for occurrences of *the N of (the) N*; this enabled me to specify the pattern more precisely and retrieve more complete data, without the dispersion and minimum frequency requirements associated with LB studies.

6.1. Overview of Results

An overview of the results obtained from the analyses of noun strings, nominalisations and *the N of (the) N* pattern in the CS and MR assignments is provided in Table 6.1. The frequency of noun strings and nominalisations is greater in the CS texts than in the MR assignments, while there are more *the N of (the) N* patterns in the MR assignments. These results offer a quick glimpse into the variation in the linguistic devices used to achieve density in the two genres. The challenges facing engineering students who need to be able to switch registers when attempting different types of coursework is highlighted in this brief overview of the results.

	Percentage of tokens in *the N of (the) N* pattern	**Frequency of noun strings (pmw)**	**Frequency of nominalisations (pmw)**
CS	**6%**	**NN: 72,788** **NNN: 15,861** **NNNN: 4,382** **NNNNN:1,386**	**37,481**
MR	**7%**	**NN: 61,078** **NNN: 11,174** **NNNN: 2,428** **NNNNN: 967**	**27,143**

Table 6. 1. Overview of results: the linguistic devices used to achieve density in the CS and MR assignments

Section 6.2 examines the linguistic features associated with Phrasal Complexity and Section 6.3 studies the noun types (Section 6.3.1) and noun sequences (Section 6.3.2) within *the N of (the) N* pattern in the CS and MR genres.

6.2. Phrasal Complexity

This first section on phrasal complexity will compare the use of nominalisations and noun strings and also briefly discuss attributive adjectives in the CS and MR assignments.

6.2.1. Nominalisations

Nominalization is a linguistic device which has received a considerable amount of research attention due to its prevalence in academic writing (Biber, Gray and Poonpon 2011, Halliday and Matthiessen 2004, Parkinson and Musgrave 2014, Staples et al. 2016). Systemic Functional linguists describe the formation of nominalisations as occurring when a process or quality which is congruently expressed using verbs and adjectives is re-construed metaphorically (e.g. *management* is derived from the verb *manage* and *possibility* is derived from the adjective *possible*).

Nominalisations in the MR and CS sub-corpus were retrieved using the CQL feature in Sketch Engine (see Section 3.7.1) to search for the four suffixes (*-tion, -ment, -sion, -ness*) identified by Biber (1988) as a means of automatically retrieving nominalisations. Some of the words retrieved by this means (e.g. *pigment*) were not actually nominalisations, and Biber's method also fails to retrieve nominalisations that do not end with one of these four suffixes. Nevertheless many other researchers (e.g. Gardner, Nesi and Biber 2018) have retrieved nominalisations using Biber's method, and it therefore has the advantage of facilitating comparisons of findings from different studies relating to the functions of nominalization (see Section 2.8.4). After filtering out words that had been falsely identified as nominalisations, the CS texts were found to contain about 37,481 (pmw) nominalisations while the MR texts only contained 27,143 (pmw).

6.2.1.1. Retrospective and Advance Labels

Closer examination of the examples revealed that some nominalisations function as 'retrospective labels' or 'advance labels' (Francis 1996) (see Section 2.8.5) by encapsulating, characterizing, replacing, linking, and lending cohesion to information that precedes or follows it (Aktas and Cortes 2008, Charles 2003, Flowerdew and Forest 2015, Sinclair 2004).

6.2.1.1.1. Nominalisations as Retrospective Labels in OCAW-CE

In Examples 1 and 2 below, the nominalisations *situation* and *variety* are retrospective labels (see Section 2.8.5). In Example1, *situation* refers to the information provided in the previous sentence regarding the emissions of carbon-dioxide gas in the Middle East region. It performs the function of encapsulating, linking given information to new information to carry forward the explanation about how the situation prompted the government to prioritize environmental sustainability. Francis (1996) only considered nouns to be retrospective labels if they functioned cohesively across clauses; the placement of the word *situation* clearly fits this definition as it is placed in the sentence adjacent the one it refers to. It is also preceded by the deictic *this*, a feature which is usually found in retrospective labels. It seems neutral in its characterization of the information, although shell nouns have the ability not only to compress information but also to express the stance of the writer.

(1)Therefore, Middle East countries are known for high emissions of carbon dioxide per capita (Carbon Dioxide Information Analysis Centre, 2007). However, Gulf Cooperation Council countries were aware of the importance of acting in this ***situation****; therefore, they have made environmental sustainability their highest priority…*

(2) It is characterized by different topography for example the mountains in some areas and the agriculture land in other areas. Oman has extended desert areas; this ***variety*** *gives them a great* ***opportunity*** *in the existence of copious amounts of groundwater.*

In Example 2, *variety* refers to the different types of topography of Oman, including desert, mountainous and agricultural areas. Again, *variety* encapsulates the claim made in the preceding sentence. In this instance the retrospective label also performs an additional

function of indicating to the reader how that 'stretch of discourse is to be interpreted' (Francis 1996: 85); drawing attention to the opportunities afforded by differences in topography. The characterization of the information is positive because the student uses the positive word 'opportunity' to persuade the reader of the need to exploit ground water sources in Oman. This retrospective label is also preceded by the deictic *this,* as in Example 1.

6.2.1.1.2. Nominalisations as Advance Labels in OCAW-CE

In Examples 3 and 4 below, the nominalisations *situation* and *location* are used as advance labels. In Example 3, *situation* has a predictive and organizing function. It tells the reader to expect an explanation of the 'situation', and functions as a link to the explanation that follows. The word does not ostensibly convey any stance but it does hint that the existing reality is not ideal and must be 'addressed'. However, the explanation is provided within the same clause in this instance. In the next example (Example 4), the word *locations* helps the reader to predict that a description of the two locations will follow.

(3)…to improve the ***situation*** *on late payments in the construction industry...*

(4) The major geometry problem in this Street is in two ***locations****, the first is at Sultan Qaboos University entrance and the second at Middle East College entrance.*

6.2.1.2. Nominalisations as technical words

In both genres in OCAW-CE, a small percentage of nominalisations are used as technical words (e.g. *porosity*, *elasticity*). This function of nominalisations has also been pointed out by Halliday and Matthiessen (2004) and Biber and Gray (2016). I consulted faculty from the Civil Engineering department who helped me identify the technical words used as nominalisations in the sub-corpora.

Some students tended to define or explain these technical words when they are used for the first time, and the subject lecturers clarified that this is what they expected (Examples 6, 7 and 8). The lecturers explained that they gave lower scores to students who did not define new terms. Awareness of this pedagogical requirement is clearly demonstrated in Example

5 in which the student clearly spells out to the examiner that his 'report' contains a definition of the term.

In Example 5, the technical meaning of *construction* is only conveyed when it is preceded by another non-technical attributive adjective to form the noun phrase *sustainable construction*. The subject lecturers explained to me that in order to qualify as being 'sustainable', a construction project needs to meet certain international assessment criteria. The technical concept expressed by this string is obviously inaccessible to disciplinary outsiders.

(5) *This report has discussed the **definition** of the **sustainable construction** and sustainable development.*

In Examples 6 and 7, the nominalisations *elasticity* and *plasticity* are followed by their definitions (underlined). Clearly, the nominalized word is technical in nature and the pedagogical purpose of the genre is highlighted by the inclusion of such definitions. In Example 8, the nominalization *compaction* is a technical word (bolded) and the following description of the properties of shale (underlined) help to explain its meaning.

(6) ***Elasticity** is the ability of materials to retrieve the real shape after removing the stress which applied on them.*

(7) ***Plasticity**: the ability of the material to hold its shape after removing the stress which applied on them.*

(8) *Marine shale forms from the **compaction** of silt and mud, it's made of many thin layers, and it can readily be split into thin pieces easily.*

As can be seen, Omani students express technical meaning using nominalisations and this finding concurs with the literature which discusses this characteristic of nominalisations (Halliday and Matthiessen 2004).

The purposes of using nominalisations to achieve cohesion and density, and to express scientific concepts (Halliday and Martin 1993, Halliday and Matthiesson 1999,

Halliday and Matthiessen 2004) are demonstrated through these examples. Many studies over the last decade have focused on this linguistic feature in different types of student writing (Gardner, Nesi and Biber 2018, Staples et al. 2016) (See Section 2.8.4). It is therefore important for pedagogical purposes to find out why and to what extent the Omani students use nominalisations, and also to explore generic variation in their use.

6.2.2. The use of Nominalisations in OCAW-CE

My findings clearly demonstrate that genre plays a role in writers' decisions about whether or not to nominalize. Although they are partially supported by Gardner, Nesi and Biber's finding (2018) that students' Case Studies in the Physical Sciences favour nominalisations, no other study has specifically studied genre differences in the use of this feature in the discipline of Civil Engineering.

In addition to using nominalisations more frequently, Case Study writers used many more different kinds of nominalization than Methodology Recount writers. Nearly 600 different nominalisations were used in the CS assignments while the figure is closer to 200 in the MR assignments.

6.2.3. Noun Strings

Noun strings are another linguistic feature discussed in relation to phrasal complexity (Biber, Gray and Poonpon 2011, Gardner, Nesi and Biber 2018, Staples et al. 2016). Biber and Gray (2016) note that pre-modifying nouns were originally only used in titles (e.g. *Miss, Captain*) and place names (e.g. *Greenwich Park*), and it was only in the twentieth century that there was a 'historical extension' to meaning relationships, and strings of common nouns began to appear. They report that 'the shift towards the present-day reliance on nouns as nominal pre-modifiers in academic written prose is perhaps the most dramatic quantitative-historical change in English witnessed in the last three centuries' (2016: 170). The importance of noun strings in scientific writing has been acknowledged by other prominent researchers too. For example, Montero (1996) noticed the prevalence of nominal pre-modifiers in computer science texts (see Section 2.8.7) and Gardner, Nesi and Biber (2018) reported their

prevalence in student writing in the hard sciences. In this thesis, I use the term *Noun Strings* to refer to nouns used as pre-modifiers and their accompanying Head Nouns.

I retrieved Head Nouns having up to four noun pre-modifiers from both the sub-corpora. The CS texts contain more noun strings than the MR assignments when their normalized frequencies are compared (see Figure 6.1). There is a very noticeable inverse relationship between the number of noun strings and the number of noun pre-modifiers.

These findings regarding the number and length of noun strings concur with Montero (1996), who found that, of the 4235 complex nominals in his dataset of computer science texts, the majority (about 97%) consisted of strings with only one or two noun modifiers. However, although the use of more than two noun modifiers with a single Head Noun is relatively rare, NNN and NNNN can still be found in both the OCAW-CE sub-corpora. Examples of the types of noun strings of different lengths that were retrieved from the CS and MR sub-corpora are shown in Table 6.2.

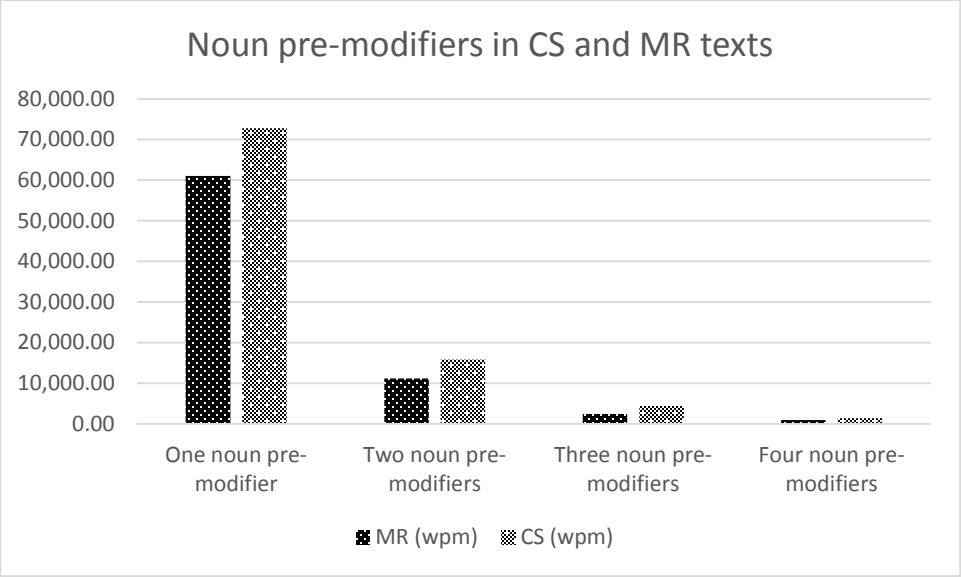

Figure 6. 1. Noun pre-modifiers in MR and CS assignments in OCAW-CE

Noun strings in the MR assignments mainly refer to the samples (e.g. *ore deposits*) and equipment (e.g. *Liquid Limit device, Measuring beaker*) used in experimental procedures,

while those in the CS texts contain proper nouns and technical terms related to the discipline (e.g. *diode lighting technology*).

Number of nominal pre-modifiers before the Head noun	**CS**	**MR**
One	*water resources; seawater desalination*	*ore deposits; sedimentary rocks*
Two	*wastewater treatment plants; diode lighting technology*	*soil particle size; liquid limit device*
Three	*Oman water conservation plan; Middle East College entrance*	*meta quartzite purple slate; steel tensile test scope*
Four	*Shaikh Hassan bin Hassan bin; Lusail Light Rail Transit System*	*can density bottle measuring beaker; soil particle size distributions test*

Table 6. 2. Examples of noun strings of various lengths retrieved from the CS and MR sub-corpora

Most noun strings in CS and MR texts achieve information density by moving modifying elements from a post-nominal to a pre-nominal position and removing prepositional markers and relative restrictive clauses, as explained by Montero (1996). In Example 9, instead of writing 'the ratio of water and cement' students have compressed this information into three words, 'water cement ratio' by removing the prepositional marker *of*. This string is also an example of common words brought together to express a technical meaning. Biber and Gray (2016) point out that removing words that signal the relationship between the nouns makes the structure more compressed but the meaning less explicit. The meaning can then only be deciphered by disciplinary insiders. Similarly, in Example 10, instead of writing 'the plants

used to purify surface water', the compressed phrase 'surface water purification plants' eliminates the restrictive relative clause and is used to express the technical concept.

(9) *The purpose of the study is to know how much of the* ***water cement ratio*** *can be used into the concrete by using* ***slump cone*** *and* ***compaction factors***.

(10) *Increase in* ***surface water purification plants***, ***wastewater treatment***, *to reduce the consumption of aquifers…*

The expression of scientific concepts by stringing together two nouns can also be seen in examples such as *slump cone, compaction factors* and *wastewater treatment*. In these noun strings non-technical words are combined in order to express technical meaning, a function of noun strings also noted by other researchers (Biber and Gray 2016, Montero 1996).

6.2.3.1. Types of noun pre-modifiers

An examination of the semantic categories of the noun pre-modifiers in the NN sequence can reveal the various functions of these nouns. To examine noun strings and nouns in the *N of (the) N* pattern, this study has adopted the eight main semantic noun categories developed by Biber (2006: 248) in his study of the T2K-SWAL corpus: Abstract, Animate, Cognitive, Concrete, Group, Place, Quantity, and Technical/Concrete. Abstract nouns are 'intangible, abstract concepts or processes', Animate nouns refer to living creatures, Cognitive nouns involve mental processes, Concrete nouns are 'inanimate objects that can be touched', Group nouns denote 'a group or institution', Place nouns refer to a location, Quantity nouns specify a quantity, amount, or duration, and Technical/Concrete nouns are 'tangible objects that are not normally perceived and/or cannot normally be touched'. The framework is unclear in certain respects (see Section 3.7.2.2 for more details). For example, the Abstract and Cognitive noun categories are not mutually exclusive because nouns such as *calculation* and *determination* belong to both.

In this section, pre-modifiers will be referred to as N1 and the Head Noun as N2.

Proper nouns as pre-modifiers

In some strings in the CS sub-corpus, proper nouns pre-modify the N2. These strings usually refer to rail and road networks (e.g. *Gulf Cooperation Council railways network*), construction projects like airports (e.g. *Jeddah Airport*), famous buildings (e.g. *Sultan Qaboos Grand Mosque construction*) and to names of organisations (e.g. *United States Green Building Council*), used to benchmark the case being studied.

MR texts do not contain any noun strings made up of proper nouns.

Tangible/concrete noun pre-modifiers

Tangible or Concrete nouns used as pre-modifiers can be found in both the MR and CS assignments, because of the nature of the discipline which deals with materials used in construction (e.g. *sand, wood*) and materials found in nature (e.g. *water, rock*). The Concrete nouns have been underlined in the examples given below. As can be seen, the modifiers have different functions; *plant, wood, sand*, *rock* and *water* are the materials out of which the N2 is made, while *sewage* and *rainwater* are the 'patients' affected by the N2 (Biber and Gray 2016).

plant materials, wood chunks, sand grains, rock samples, water molecules, sewage treatment, rainwater treatment

Intangible pre-modifiers

The intangible pre-modifiers retrieved from the CS and MR sub-corpora realise a wide range of functions and meanings as can be seen from the examples below. In the string *engineering profession*, an inanimate entity (N2) is associated with the profession identified by the N1. In *quality design*, the N1 characterizes the N2 (*design*). The pre-modifiers *risk, demand*, and *power* are affected by the processes denoted by the N2 (*management, reduction, generation*). In these three examples nominalisations are used to express 'processes or activities'; a tendency noted by Biber and Gray (2016: 176).

engineering profession, quality design, risk management, demand reduction, power generation

In the examples below nominalisations are used as noun pre-modifiers (underlined) to denote 'abstract attributes and qualities' (Biber and Gray 2016: 176) which characterize the N2s:

distribution curve, classification systems, construction materials, pollution problems

The use of nominalisations, both as pre-modifying nouns and Head Nouns, seems to be a method by which Omani students achieve phrasal complexity and increase the density of their texts.

6.2.4. Discussion on the use of Noun Strings in OCAW-CE

These findings about the greater frequency of noun strings in CS texts are somewhat unexpected, as they do not match the findings of Gardner, Nesi and Biber (2018) relating to the two types of density (see Section 2.4.5). It might have been predicted that the MR texts, being more technical in nature, would contain more noun pre-modification than the CS texts. However it should be borne in mind that many of the CS noun strings contained proper nouns, and were therefore of a rather different type to the noun strings associated with compressed procedural density.

Phrasal complexity has been associated with advanced university level writing in the sciences (Gardner, Nesi and Biber 2018, Staples et al. 2016), but more advanced linguistic skills are required to convert the common nouns in clauses into noun strings, the strings of proper nouns denoting places and companies in the CS assignments should be easy for even non-proficient writers to produce.

6.2.5. Attributive Adjectives

Attributive adjectives as noun pre-modifiers have been studied as a part of phrasal complexity (Biber, Gray and Poonpon 2011, Gardner, Nesi and Biber 2018, Parkinson and Musgrave 2014). Therefore this feature was also analysed in MR and CS texts to discover the extent to which Omani university students use this feature and also if there are genre variations in its use. Comparison of the normalized frequencies show that the CS texts use

more attributive adjectives than the MR assignments (48,702 as against 41,918 wpm). The descriptive nature of the CS genre may account for this difference.

There are very few instances of limited and exaggerated adjectives (e.g. *good, amazing, huge, big, enormous*) which Ädel & Erman (2012) consider a feature of less proficient writing. Examples 11 to 14 demonstrate the use of such adjectives by the Omani students.

*(11) There are not any **huge columns** in the building, which is usually found in buildings.*

*(12) The Middle East has a **huge** scarcity of water on the surface and other natural reserves as well as less rainfall.*

*(13) The question now is, the face of the **huge** cost of the project…*

*(14) …during the visit to the project site, which is one of the **big** projects in terms of specification of building material used.*

Some studies (Biber, Gray and Poonpon 2011, Parkinson and Musgrave 2014) have tried to examine phrasal complexity features such as attributive adjectives in terms of the order of their acquisition. Attributive adjectives have been placed in a certain position in the 'developmental index' of acquisition as a feature that learners acquire before they start using pre-modifying nouns. For example, Parkinson and Musgrave (2014), in their study of assignments written by students belonging to two different levels of proficiency, found that attributive adjectives were used more prolifically by the less advanced group. Once again, however, the influence of genre was not considered in this study. In my study attributive adjectives were found to occur more often in CS texts (e.g. *economic deficit, global architectural styles, sustainable building*) than in MR texts, although there was no difference in the proficiency level of the writers of these two genres. This suggests that genre differences could have contributed to the variation in the frequency of attributive adjectives in the two sub-corpora Parkinson and Musgrave (2014) examined, although it is certainly possible that the type and range of attributive adjectives might vary with increasing levels of proficiency.

Staples et al. (2016) conducted a much more nuanced study by also considering genre and disciplinary differences in student writing. They report the increased frequency of attributive adjectives in Case Studies in the Physical Sciences, in accordance with my own findings.

Comparing these findings to those of Gardner, Nesi and Biber (2018), the use of attributive adjectives in the CS assignments continues to strengthen the notion of two types of density in advanced student writing. With the exception of noun pre-modifiers, the CS assignments exhibit the linguistic features of the second type of density while the MR assignments exhibit the linguistic features of the first type of density. These findings regarding similarities and differences in the frequency and function of noun strings, attributive adjectives, and nominalisations in CS and MR assignments have addressed RQ 3a. Genre variation is also apparent in realizations of the pattern *the N of (the) N*, discussed in Section 6.3.

6.3. Nouns in *the N of (the) N* pattern

The pattern *the N of (the) N* constitutes at least 7% of the total number of words in the MR sub-corpus and 6% of the total number of words in the CS sub-corpus, suggesting that the MR texts use this pattern slightly more than the CS texts. Considering that all the instances of the pattern may not have been retrieved (see Section 3.7.2), this percentage of the total word count is quite considerable. The LB analysis in Chapter 5 also reflects this; instances of LBs realising the pattern, *the N of (the)* (e.g. *the principles of sustainability, the top of the*), were more frequent than other LB patterns in both sub-corpora. LBs in this pattern constitute 2.98% of the tokens in the MR texts as against 0.79% in the CS texts.

6.3.1. Analysis of noun types used by MR and CS texts

To investigate the similarities and differences in the noun types preferred by the CS and MR genres, I examined the semantic noun categories used in the pattern *the N of (the) N*. I categorized the first noun in the strings retrieved from the CS and MR sub-corpora according to the eight noun types identified by Biber (2006) and described in Section 3.7.2.2: Abstract, Animate, Cognitive, Concrete, Group, Place, Quantity, and Technical/Concrete. As I also

found proper nouns and gerunds in *the N of (the) N* strings retrieved from the two sub-corpora, I added two additional categories: Proper Noun and Gerund. I adopted Huddleston and Pullum's (2002) definition of Gerunds as nouns which are formed from verbs with the addition of the suffix *–ing*. These 'gerundial nouns' perform the functions of nouns by taking modifiers and performing as the subjects and objects of clauses.

Figure 6.2 shows the distribution of the ten noun categories in the first position in the string (hereafter N1). A comparison of the normalised frequencies of the noun categories indicates two main differences between the two genres; MR texts use many more Quantity, Place and Concrete nouns than the CS texts while the CS texts use Abstract nouns much more frequently. Secondly, the CS texts contain three noun categories, Group, Animate and Proper, which are not represented at all in the MR sub-corpus. This preliminary comparison highlights genre differences in the use of noun categories.

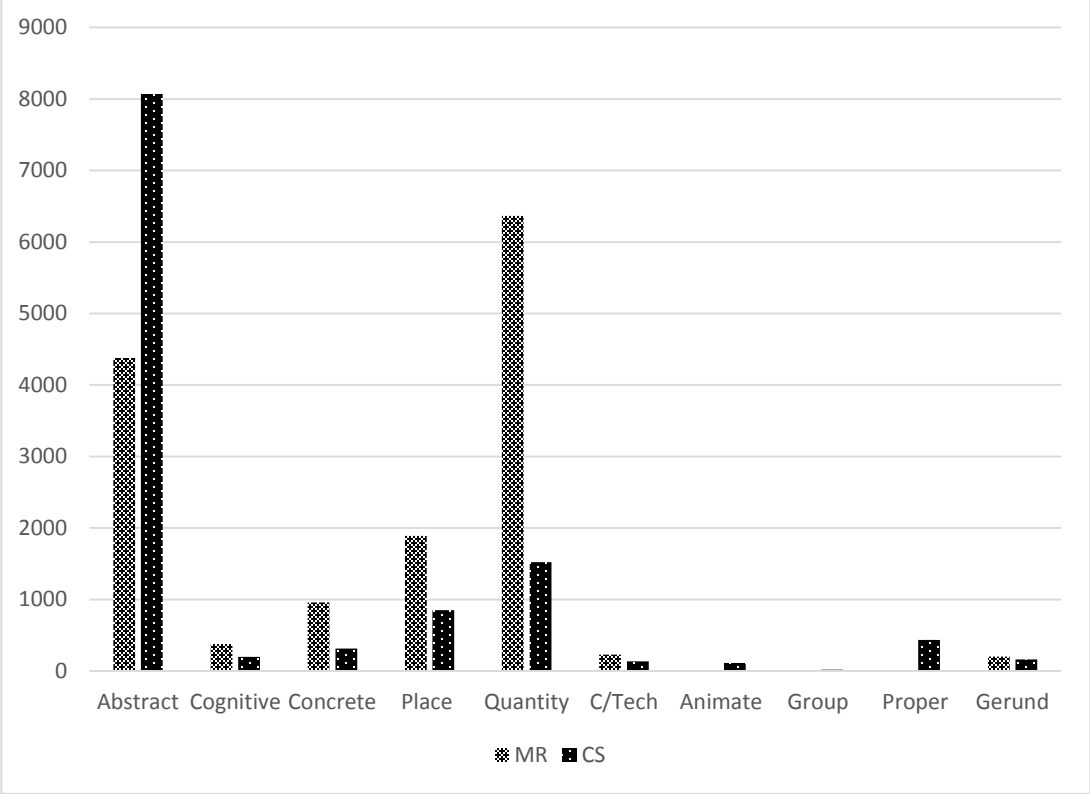

Figure 6. 2. Semantic categories of N1 in the MR and CS sub-corpora (normalised frequencies)

I applied the z-test to test for statistically significant differences in the means of noun categories. This test was adopted because these sample sizes (31 CS texts and 63 MR texts) exceeded 30 and are therefore considered 'large' in statistical terms (Blaikie 2003). Standard deviations were calculated by considering the distance of each data point (in this case, the frequency of a noun category or noun sequence in each text) from the mean. This method takes into account the fact that certain noun categories might be present in large numbers in some assignments while they are rare or absent in others; it thus reduces the possibility of distortion due to idiosyncratic differences in writing styles or the nature of tasks within the same genre.

I recorded the raw frequency of each of the ten noun types (N1) and normed their occurrence in each text (per 1000 words) before applying the z test. The results are provided in Table 6.3. The confidence interval was set at 95% and a comparison of the means of the different noun categories in the CS and MR assignments shows that the differences are in all probability not due to chance (Greenland et al. 2016). Four categories do not show any significant difference: Cognitive, Concrete/Technical, Group and Gerund.

S. No	***N 1***	***Mean (MR)***	***Mean (CS)***	***sd (MR)***	***sd (CS)***	***Z Value***	
1	Abstract	3.698	8.275	3.453	3.863	5.589	**
2	Cognitive	0.369	0.211	1.198	0.395	0.947	ns
3	Concrete	1.004	0.256	1.577	0.434	3.506	**
4	Place	2.020	0.811	3.601	0.702	2.568	*
5	Quantity	7.042	1.520	5.837	1.172	7.219	**
6	Concrete-Technical	0.191	0.177	0.562	0.365	0.146	ns
7	Animate	0.000	0.114	0.000	0.214	2.959	**
8	Group	0.000	0.024	0.000	0.094	1.437	ns
9	Proper	0.000	0.411	0.000	0.653	3.503	**
10	Gerund	0.183	0.166	0.621	0.342	0.180	ns

Table 6. 3. Frequencies of N1 categories in the CS and MR sub-corpora

*** Indicates significance at 1% level*

** Indicates significance at 5% level*

***ns** indicates that the differences are not significant (p > 0.05)*

The distribution of each of these N1 noun categories and the differences and similarities in their use in the two genres are discussed in the following sections.

6.3.1.1. Abstract Nouns

Abstract nouns as N1s in the pattern *the N of (the) N* are used in both MR and CS genres although to a much greater degree in the CS texts. The normalised frequency of Abstract nouns in the CS texts (e.g. *aspirations, dissemination, expectations)* is about 8071 pmw, which is double that in the MR texts (e.g. *change, delay, design*).

On applying the z-test, I found that there is a significant difference ($p<0.05$) in the scores obtained for the abstract noun category between MR (M=3.69, SD=3.45) and CS (M=3.86, SD=3.86) texts; z =5.58.

In the CS sub-corpus, Abstract nouns that are rare in general corpora occur with quite a high frequency. For example, the frequency of *sustainability* is 52 pmw in the 15 billion-word English Web 2015 corpus while in the CS sub-corpus it is about 845 pmw. Other examples of Abstract nouns which are relatively rare in this web corpus (less than 75 pmw) but which are used quite frequently in the CS (more than 100 pmw) include *conservation, estimation, framework* and *installation*. Technical words that are Abstract nouns are rare in both sub-corpora, although there are a few exceptions (e.g. *plasticity, porosity*) in the MR texts (see 5.3.1 for more details on Nominalisations used to express technical meaning in noun strings).

Many studies have reported the use of abstract nouns in academic texts (e.g. Biber 2006, Gardner, Nesi and Biber 2018). Biber (2006) found that abstract nouns constituted 50% of the noun types in written university registers. Interestingly, although 50% of the N1s used in *the N of (the) N* pattern in the MR texts are Abstract, in the CS texts more than 70% are Abstract. These findings strongly suggest that genre variation may account for noun category preferences.

6.3.1.2. Quantity Nouns

Comparing the normalised frequencies, the MR texts contain four times as many Quantity nouns (6365 pmw) as CS texts (1521 pmw).

The z-scores (z=7.21) revealed a significant difference in the scores obtained for Quantity nouns in the MR sub-corpus (Mean=7.04, SD=5.83) and the CS sub-corpus (M=1.52, SD=1.17).

Most of the Quantity nouns in N1 position in this pattern in the MR texts are associated with a Concrete noun in N2 position. In Examples 15 and 16, the Quantity nouns *mass* and *weight* refer to the Concrete nouns, *cylinder* and *soil*.

*(15) Determine the **mass** of the cylinder + sand (M1) to the nearest gram.*

*(16) In sand replacement method, a small cylindrical pit is excavated and the **weight** of the soil excavated from the pit is measured.*

Even though Quantity nouns are used in both sub-corpora, they are used for different purposes. The normalised frequencies of the nouns *mass, weight* and *volume* are much higher in the MR texts than in the CS texts, as Table 6.4 demonstrates. The examples in Table 6.5 illustrate the use of these nouns in MR texts as part of the recount of experimental procedure involving the measurement of volume, mass and weight of concrete objects which are either samples (e.g. *soil, sand*) or lab equipment (e.g. *calibrating cylinder, beaker*).

	CS (pmw)	MR (pmw)
Mass	56.39	2,901.59
Weight	217.49	3,909.95
Volume	185.27	2,839.86

Table 6. 4. Normalised frequencies of mass, weight and volume in the CS and MR sub-corpora

Some of these nouns in the MR assignments occur in typical patterns of use. Table 6.5 shows how the phrases *the mass of the, the weight of the,* and *the volume of the* are repeatedly used with different concrete nouns.

Quantity noun (N1) in *the N of (the) N* pattern	**Second noun in *the N of (the) N* pattern (N2)**
the mass of the	calibrating cylinder excavated soil water sand pouring cylinder pan given object
the weight of the	beaker sieve sand pouring cylinder samples liquid object soil sand
the volume of the	container calibrating cylinder given beaker given cube sample

Table 6. 5. Noun strings with the pattern the Quantity Noun of the Concrete Noun in MR assignments

In the CS texts, on the other hand, there are references to time periods (e.g. *date, duration, period*), followed by a reference to the time required to complete an activity. This is perhaps indicative of genre difference, reflecting the fact that CS assignments are usually concerned with the duration of a construction project or the date of its completion. Example 17 uses time reference (*duration*) to specify the time taken to complete the first phase of construction of a project.

*(17) The **duration** of the first phase is 2009 to 2016, the total length of the transit network is 24 km and the transit technologies…*

Some of the same Quantity N1s are found in the pattern *the N of (the) N* in both the MR texts and the CS texts (e.g. *volume*, *size*, *area*). However in the CS texts they are followed not only by Concrete nouns but also by other noun categories such as Abstract (e.g. *the volume of traffic*), Place (e.g. *the size of the bedroom*) and Proper nouns (e.g. *the area of Qatar*). The nature of the CS genre which involves explanations of the case and its context

may account for the use of these different noun categories in the second position in *the N of (the) N* pattern. For example, when a stretch of highway is taken as the case, details about many factors such as the volume of the traffic and the peak hours are included in the explanation. On the other hand, the MR assignments are restricted in their scope because they focus on experiments and the Concrete objects used to conduct experiments.

6.3.1.3. Place Nouns

The normalised frequency of Place nouns as N1s is 1886 pmw in the MR texts and 850 pmw in the CS texts. There is a statistically significant difference in the means in the MR texts (M=2.02, SD=3.6) and the CS texts (M=0.81, SD=0.7) with $z=2.56$, $p<.05$.

All the N1 Place nouns (bolded) in the MR texts (except one, *environment*) refer to the position or location of concrete objects (underlined) as illustrated in Examples 18 and 19. MR Place nouns also usually refer to relatively smaller concrete nouns (e.g. *cup, sieve, cylinder, stopper*), usually a piece of laboratory equipment. Sometimes the second Concrete noun that follows the Place noun in the pattern is a sample used as part of the experiment being recounted.

*(18) The **base** of the cup is filled with soil and a groove is then made through the soil to the base of the cup.*

*(19) A thin layer on the **outside** of the nodules is usually different in color, typically white and rough in texture.*

The Place nouns in the CS texts mainly refer to the position of objects (e.g. *the **front** of the tractor, the **surface** of the door jamb*), to cities and countries (e.g. *the South of Shanghai, the city of Mumbai*) or direction (e.g. *south, northwest*). This is because case studies in the OCAWE-CE corpus usually involve recounting some background about the case with reference to its location in a city or country. In other cases, position is indicated with reference to large objects or phenomena (e.g. *building, lake, wall, tower*). Example 20 points to the location of a rail network, which is the case being studied, while Example 21 refers to the location of something on a relatively large object (wall).

*(20)…new rail network in UAE, this network an underground line is start to Dubai Airport with the port and indusial area of Jabel Ali in the **south** of the emirate.*

*(21)…removal of the soil in the **surface** of the wall until reach the steel.*

6.3.1.4 Concrete Nouns

The comparison of the normalised frequencies reveals that Concrete nouns are three times more frequent as N1s in the pattern *the N of (the) N* in MR texts (about 956 wpm) than in CS texts (about 312 wpm). The z-test for significance revealed that there is a significant difference in the means for the MR texts (M=1, SD=1.57) and the CS texts (M=.25, SD=.43), with $z = 3.5$, $p < .05$.

Like the Place nouns, the Concrete N1s in the strings retrieved from the CS sub-corpus mostly refer to larger objects (*bridge, sleepers*) as Examples 22 and 23 show. Examples 24 and 25 serve to illustrate the use of concrete nouns in the MR sub-corpus referring to parts of laboratory equipment (*nozzle, hook*) used to carry out the experiments described in Methods sections. Examples 22 to 25 also show *the N of (the) N* patterns where both the N1 and the N2 are Concrete nouns, with the second noun naming a piece of equipment used in an experiment.

(22) … *the most successful section in the design of building are a street and the **bridges** of glass over it.*

(23) *Railway ballast is the foundation of material placed and packed around and below the **sleepers** of railway track.*

(24) *Dropped the cube in the water and the water started flowing in the beaker through the **nozzle** of Eureka Can.*

(25) *Attach a string to the given object and suspend it from the **hook** of the spring balance until it completely submerge in the Eureka Can without touching the bottom or sides of the Can.*

Biber (2006) notes that concrete nouns are common in engineering textbooks. In the CS and MR sub-corpora, however, Concrete nouns are less frequent than other noun categories in the pattern *the N of (the) N*.

6.3.1.5. Cognitive Nouns

The number of N1 Cognitive nouns in the MR sub-corpus is higher than the CS texts in terms of normalised frequency (377 pmw vs. 198 pmw), but there was no significant difference between their frequency in the CS texts (M=.211, SD=.395) and the MR texts (M=.369, SD=1.198) texts with z =.947, p<.05.

Six types of Cognitive nouns are used as N1s in the CS sub-corpus (*idea, concept, view, calculation, diagnosis, analysis*), while only four types are used in the MR sub-corpus (*calculation, determination, analysis, theory*). The noun *calculation* is repeated eight times and *determination* is repeated four times in the MR texts, mainly to explain the theory underpinning the experiment recounted in the assignment or to provide the scope of the experiment or 'test', as Examples 26 and 27 demonstrate.

*(26)This property is necessary for the **calculation** of seepage through earth dams or under sheet pile walls, and the **calculation** of the seepage rate from waste storage facilities*

*(27) This test method covers the **determination** of the compressive strength of cement mortars, using 50mm cube specimens.*

The types of N2 which the first Cognitive noun refers to varies between the two genres. In the CS texts, the N2 usually refers to people or concepts (e.g. *the views of others, the concept of sustainability*) while in the MR texts it mainly refers to phenomena that need to be measured (e.g. *the **calculation** of the seepage rate, the determination of the particle size distribution*).

6.3.1.6. Concrete-Technical Nouns

A comparison of normalised frequencies indicates that the MR texts contain a much higher frequency of Concrete-Technical Nouns as N1s (226 wpm) than the CS texts (132 wpm).

However there was no significant difference between the MR texts (M=.19, SD=.56) and the CS texts (M=.17, SD=.36) with z=.14, p>.05.

Biber (2006) reports that the Concrete-Technical nouns found in the T2K-SWAL corpus are highly specialised even though they are frequent. In the CS and MR texts too, they often refer to discipline-specific concepts. For example, the word *bill* may sound non-technical but is part of the technical phrase *bill of quantities* in the area of Quantity Surveying in Civil Engineering (Example 28). Similarly, the words *axis* and *line* (Examples 29 and 30) are used as parts of technical terms which might only be meaningful to disciplinary insiders.

*(28) This document used to full any extra work doing in the construction project. The extra work not includes in the **bill** of quantities.*

*(29) In the crushing test, care had taken to ensure that the cubes had properly positioned and aligned with the **axis** of the thrust of the compression machine to ensure uniform loading…*

6.3.1.7. Gerund

Although the category of nouns which I have termed Gerunds only accounts for about 2% of the nouns in the pattern *the N of (the) N* in the CS and MR texts, I counted them as a separate category because they occur in sufficient number, and also because they do not find a place in Biber's (2006) semantic noun category framework. When the normalised frequencies of Gerunds in the pattern in the two sub-corpora were compared, it was found that they were more frequent in the MR texts (201 wpm) than in the CS texts (161 wpm). The results of the z-test revealed that this difference is not significant (MR texts (M=.18, SD=.62) and CS texts (M=.16, SD=.34) with z=.18, p>.05).

Gerunds are used in CS and MR assignments to refer to common processes and activities in Civil Engineering such as *planning, paving, painting,* and *lighting*. Gerunds have all the characteristics of nouns: in Examples 30 and 31, the gerunds *monitoring* and *opening* are modified by prepositional phrases (underlined). In Example 30, *monitoring* is also modified by the attributive adjective *efficient*.

(30) Additionally, due to the highly efficient ***monitoring*** *of energy of Green-Nest, it is certified as platinum.*

(31) On the date of May 15, 1889 was the ***opening*** *of the Eiffel Tower.*

6.3.1.8. Animate, Group and Proper Nouns

Animate, Group and Proper Nouns are not used as N1s in the pattern *the N of (the) N* in the MR sub-corpus and there are only small percentages of these nouns in the CS sub-corpus. Since MR assignments are tasks involving the recounting of experiments where the procedure rather than the doer is foregrounded, the absence of Animate nouns is not surprising. The CS assignments, on the other hand, involve descriptions of the background and history of cases where the doer of the action is important. In Examples 32 and 33, the *designer* of a building and the *mayor* of a town are mentioned, and subject lecturers informed me that the nature of the CS genre calls for this kind of detail. In Examples 34 and 35, the use of Group nouns in CS texts is demonstrated; the writers provide details (*the Gulf Cooperation Council, Research Council of Oman*) as part of background before the case is analysed.

(32) *Called the tower that name in relation to the* ***designer*** *of Burj engineer Gustave Eiffel, the company that designed and built the tower founder.*

(33) *The building design still needs to be approved by the* ***mayor*** *of Wuhan, but the construction could start as soon as …*

(34) *The* ***members*** *of* ***Gulf Co-operation Council*** *are Kuwait, Bahrain, Qatar, Oman, Saudi Arabia, and the United Arab Emirates.*

(35) *HCT Green-Nest is the first eco house design competition that the* ***Research Council of Oman*** *has sponsored.*

Proper nouns constitute at least 4% of the nouns in the pattern *the N of (the) N* in the CS texts, but are absent in the MR texts. Proper nouns in the CS genre are mainly place

names, as in Example 36, since many of the CS assignment briefs (see Appendix 3) require writers to make a comparison between two countries or locations.

(36) …*the various underlying factors with relation to water conservation and management in the* ***Kingdom of Bahrain*** *and also in the country of* ***Sultanate of Oman****.*

6.3.1.9. Range of nouns in the first position of *the N of (the) N* pattern

The discussion of these noun categories can be completed by summarising the range of the different N1 types in each category in the CS and MR assignments. This analysis will help us estimate the quantity of noun types that a student might need to be familiar with in order to use the pattern *the N of (the) N* in these genres.

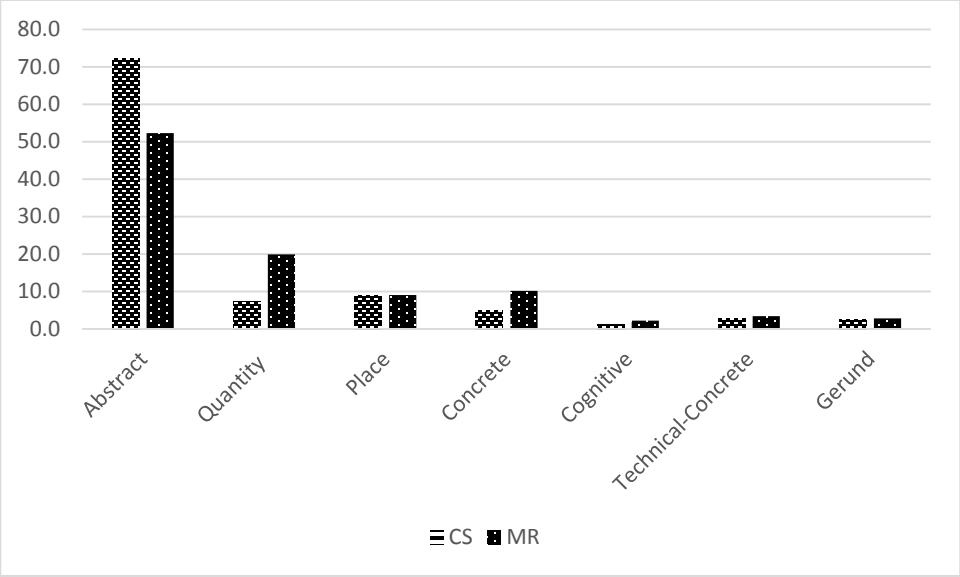

Figure 6. 3. Percentage of N1 categories in the pattern the N of (the) N

Figure 6.3 shows the percentage of N1s in each of the categories in the two genres. The biggest difference between the CS and MR assignments is in three categories: Abstract, Quantity and Concrete Nouns. We can tentatively conclude that since there is a greater range of Abstract N1s in the CS texts than in the MR texts, students will need to be familiar with more Abstract noun types when attempting the CS genre. The Case Studies in Civil Engineering involve many components, including an explanation of the theory and background to the issue, a review of the literature, an evaluation and recommendations.

Quantity N1 types are much higher in the MR assignments than in the CS assignments; this difference could be because of the nature of the MR genre involving experimental procedures with calculations and measurements. The frequency of Concrete N1 types is also much higher in the MR sub-corpus. This increased frequency can be ascribed to the use of concrete objects as samples (e.g. *soil, sand, rock*) and the various pieces of equipment used in experiments (e.g. *crank, cup, cone, shutter, particles*).

Section 6.3 has discussed the differences between the CS and MR assignments in their use of N1 types in the pattern *the N of (the) N*. These differences can be seen as distinct genre traits since the groups of writers of both sub-corpora belong to the same writing culture, language background, and discipline.

The next section will continue the discussion of the types of nouns in the pattern *the N of (the) N* by identifying the sequences of noun types.

6.3.2. Noun sequences in *the N of (the) N* pattern in CS and MR assignments

Within the pattern *the N of (the) N*, there are 27 types of noun sequences in the MR assignments and 55 types in the CS assignments; 25 of these are common to both genres. Figure 6.4 illustrates the percentages of N1-N2 sequences: as can be seen, some of them are very rare. For example, the Concrete-Quantity sequence is represented in only one of the 63 texts in the CS sub-corpus and there is not a single instance of it in the MR sub-corpus.

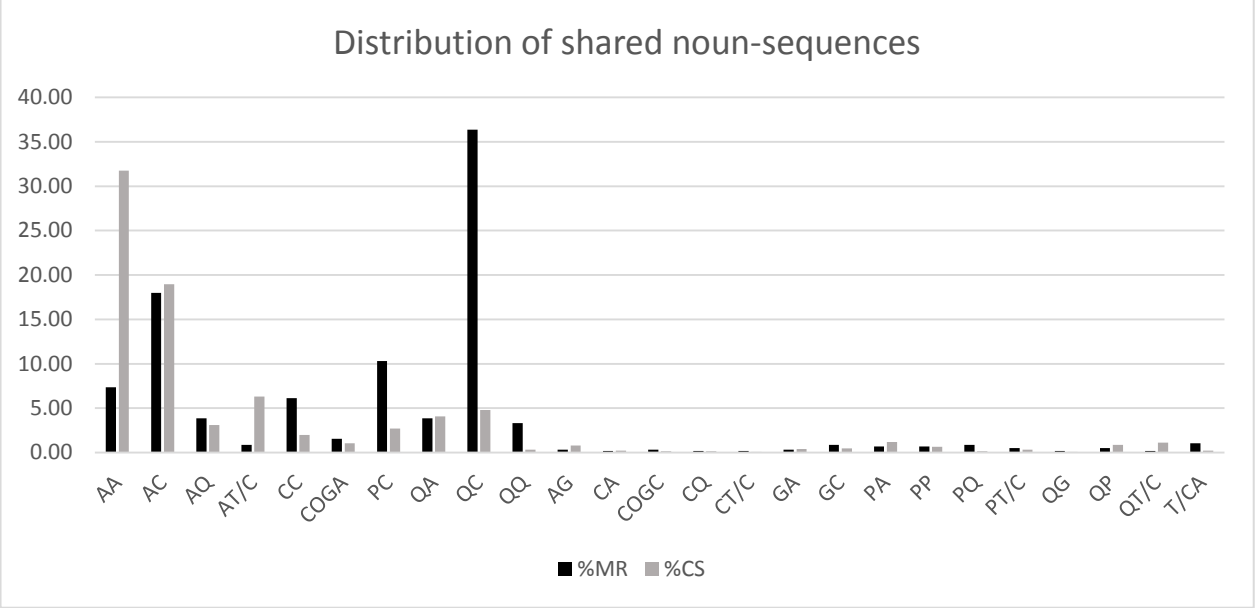

Figure 6. 4. Distribution of shared noun-sequences in MR and CS assignments in the N of (the) N pattern

A z-test was applied to investigate whether the differences in the means of these patterns is statistically significant. The normalized counts (per 1000 words) for each noun sequence category was obtained before applying the z test. The six noun sequences showing significant differences in frequency are tabulated in Table 6.6: Abstract-Abstract, Place-Concrete, Quantity-Concrete, Abstract-Technical/Concrete, Concrete-Concrete and Quantity-Quantity. Examples of these types of noun sequences are provided in Table 6.6. The first noun in the Noun-Noun sequence has been bolded and the second noun in the Noun-Noun sequence has been underlined to highlight the types of nouns that are found frequently together in *the N of (the) N* sequences in MR and CS assignments. The distribution of N1-N2 sequences by percentage is provided in Figure 6.4.

	Mean (MR)	*Mean (CS)*	*sd (MR)*	*sd (CS)*	*z value*	*significance*
Abstract-Abstract	3.834	0.926	2.145	1.521	6.760	**
Place-Concrete	0.316	1.504	0.422	2.821	3.268	**
Quantity-Concrete	0.533	5.812	0.500	5.564	7.469	**
Abstract-Tech/Concrete	0.802	0.131	0.764	0.505	4.433	**
Concrete-Concrete	0.175	0.941	0.357	1.566	3.692	**
Quantity-Quantity	0.161	0.474	0.289	1.142	2.046	*

*** Indicates significance at 1% level*

** Indicates significance at 5% level*

***ns** indicates that the differences are not significant (p > 0.05)*

Figure 6. 5. Frequencies of noun sequences in CS and MR texts

	MR	CS
Abstract-Abstract	*the **principles** of mechanics; the **consistency** of the compaction*	*the **objective** of the project; the **quality** of the air*
Place--Concrete	*the **base** of the dam; the **outside** of the bottle*	*the **interior** of the building; the **surface** of the wall*
Quantity-Concrete	*the **mass** of an object; the **volume** of water*	*the **width** of the lane; the **amount** of refrigerant*
Abstract-Technical/Concrete	*the **type** of the graphs; the **types** of project*	*the **life** of the project; the **need** for artificial light*
Concrete-Concrete	*the **cone** of the cylinder; the **hook** of the spring balance*	*the **sleepers** of railway track; the **roof** of Malaysian houses*
Quantity-Quantity	*the **percentage** of the total weight; the **ratio** of the total mass*	*the **period** of time; the **ratio** of the weight*

Figure 6. 6. Examples of noun sequences with significantly different frequencies in CS and MR texts

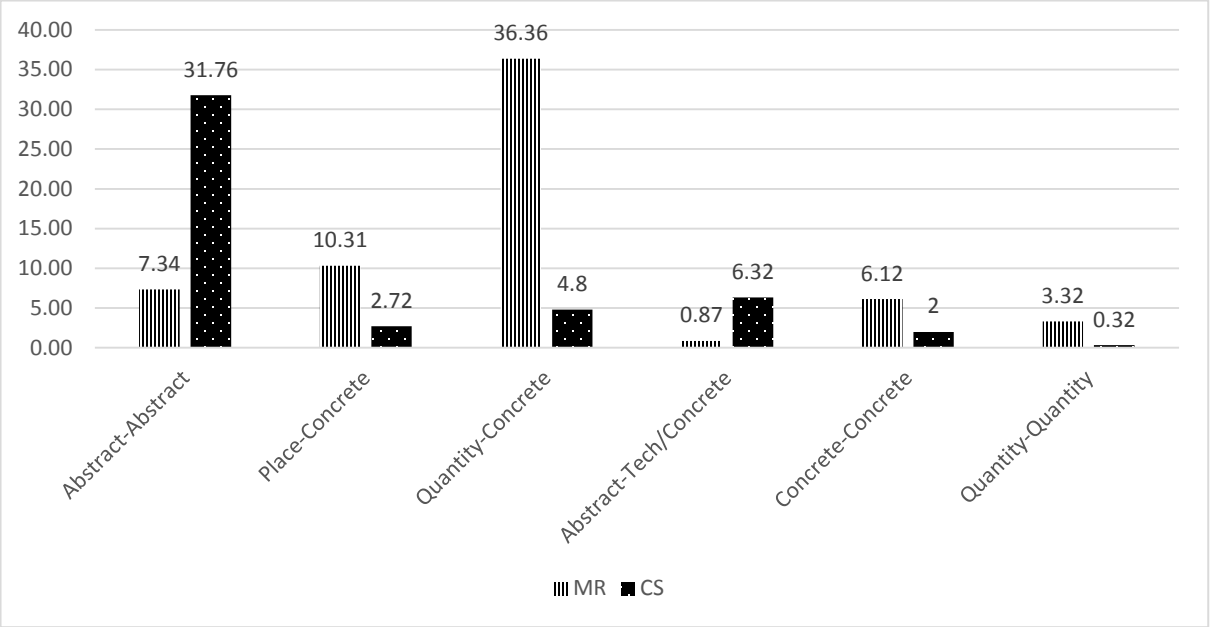

Figure 6. 7. Distribution of noun sequences in *the N of (the) N* pattern in MR and CS assignments

The semantic categories of N1-N2sequences with statistically significant differences in their means are Abstract, Place, Concrete, Quantity and Technical/Concrete. In terms of percentage, N1-N2 sequences containing Abstract nouns are found more in the CS texts and N1-N2 sequences containing Concrete, Quantity and Place nouns are found more in the MR assignments. These findings reflect those of the first analysis (see Section 6.3) where I report that Abstract nouns are preferred in the CS texts while Place, Quantity and Concrete nouns are preferred in the MR texts. The noun sequences showing a much higher percentage in the MR texts belong to the Quantity-Concrete, Place-Concrete, Concrete-Concrete, and Quantity-Quantity categories (see Section 6.3.2 for examples). This finding is hardly surprising as it has already been seen (Section 6.3.1.2) that these noun sequences are used in MR texts because of the frequent need to refer to the quantity, location and parts of the samples, or objects which are used in experiments.

6.3.3. Discussion of the noun categories in *the N of (the) N*

The differences in the noun categories used in the pattern *the N of (the) N* both in terms of N1 position and in terms of N1-N2 sequencing indicate preferences based on genre. Clearly

the pattern is used to achieve density in both genres as it accounts for at least 6% and 7% of the total words in the CS and MR sub-corpora.

Statistically significant differences were found in the means of six of the semantic noun categories in the pattern *the N of (the) N* in the CS and MR sub-corpora: Abstract, Quantity, Concrete, Place, Animate and Proper. Having investigated the differences in the means, the normalized frequencies of each of the semantic noun categories were compared between the two sub-corpora. This comparison indicated the preference of each sub-corpus for a specific noun category. For example, Abstract nouns are more frequently used in CS assignments while Quantity, Concrete and Place Nouns are preferred in the MR assignments. These findings address RQ 3b by identifying the types of nouns found in *the N of (the) N* pattern in CS and MR assignments, and the functions they perform within this frame.

Both Abstract N1s and the N1-N2 Abstract-Abstract sequence are preferred in CS assignments. The less frequent use of Abstract nouns and the greater use of Quantity-Concrete N1-N2 sequences in the MR assignments indicate 'compressed procedural density', as identified by Gardner, Nesi and Biber (2018). They report that MR texts in the Physical Sciences in the BAWE corpus are characterised by this first type of density. Based on these findings that Abstract nouns, nominalisations and attributive adjectives are more prevalent within the pattern *the N of (the) N* in CS assignments, we might tentatively conclude that the linguistic devices used by the CS texts tend towards the second type of information density identified by Gardner, Nesi and Biber (2018).

These findings therefore indicate that a more nuanced understanding of phrasal complexity, taking genre and other situational variables into account, is required to identify the linguistic features used to achieve density. For example, the use of nominalisations or attributive adjectives may not be indications of the level of writing proficiency but rather the traits of a genre. Researchers such as Biber, Gray and Poonpon (2011, and Parkinson and Musgrave (2014) need to re-examine their corpora and consider the differences between genres within the same discipline before generalizing about the characteristics of proficient writing. Studies that have compared expert and student writing (Chen and Baker 2010,

Hyland 2008) also need to be wary of drawing firm conclusions from the results of analysis of corpora containing genres that are not comparable to pedagogical genres.

6.4. Conclusion

This chapter has thus attempted to answer RQ 3 regarding variation between CS and MR assignments. It has sought to uncover variation in the linguistic devices used to achieve density in the CS and MR genres by investigating phrasal density features, *the N of (the) N* pattern, and the semantic noun categories used within this pattern in the two sub-corpora.

The investigations of noun categories in the two sub-corpora revealed that two noun categories may possibly be added to Biber's (2006) framework to better accommodate pedagogic genres such as CS and MR assignments. My analyses indicate that frameworks designed on the basis of corpora which are not representative of student assignments, such as that used by Biber (2006), may need to be adapted as I have done in my analysis. The findings regarding genre variation in preferred noun categories can be explored further using larger corpora if programs are developed to automatically tag for noun categories.

The findings in the previous sections on phrasal complexity and the noun categories used in *the N of (the) N* pattern can be interpreted in the light of recent developments in the literature, especially that of Gardner, Nesi and Biber (2018). These findings further confirm the theory that there is more than one type of density, achieved through the use of different linguistic features. The MR texts tend to favour compressed procedural density while the CS assignments tend to favour information density (see Section 2.4.6). These analyses highlight the challenges faced by engineering students who need to switch between the two types of density as they attempt these two genres.

Chapter 7

Conclusion

7.1. Aim of the study

The aim of this thesis was to attempt a linguistic description of the two main genres, Case Studies and Methodology Recounts, in the Omani Corpus of Academic Writing-Civil Engineering (OCAW-CE). The importance of writing in Civil Engineering (CE), both in academic and workplace contexts, has already been discussed (see Chapter 1). Clarity in written communication is a crucial factor in CE because of the liability issues that can arise due to ambiguous wording of texts (Conrad 2017) and also because of safety factors.

Little research attention has been directed towards the pedagogical genres in specific disciplines (Hardy and Friginal 2016), especially in the 'expanding circle' of countries where English is the medium of higher education and the lingua franca in the workplace. This is in spite of the fact that many countries belonging to such contexts, including the Gulf Cooperation Council countries, have adopted English as the language of tertiary education. In fact, there is only one systematically collected and publicly archived student assignment corpus in the world, namely BAWE (see Section 2.3.4). The challenges associated with collecting student assignments and the paucity of publicly accessible and relevant corpora may account for this dearth of research. OCAW-CE will therefore be one of the first corpora of student writing collected from one of these countries in the 'expanding circle'. This study has addressed a critical gap in disciplinary writing research by investigating coursework genres in this corpus.

I have identified the two main genres assigned within Civil Engineering through the systematic compilation and then investigation of the corpus, and examined their rhetorical structure, the phraseology and the phrasal complexity features used to realise their communicative intent.

The Research Questions posed for this purpose were the following:

Research Questions

1. What is the move structure of the two main genres in Civil Engineering?

2. What are their similarities and differences in terms of phraseology?

3. What are their similarities and differences in terms of phrasal complexity?

7.2. Summary of the study

The two main Civil Engineering genres produced by undergraduates at the Omani institution, Case Studies and Methodology Recounts, were analysed in terms of their move structure, and the linguistic devices used to realise the moves and steps were identified and described (see Sections 4.2 and 4.4). Other researchers have analysed the Lab Report genre and my analysis was guided in part by these previous attempts (Nesi and Gardner 2012, Musgrave 2017). A move analysis of the Case Study genre in Civil Engineering has not been done before although Conrad et al. (2016) did attempt to analyse the rhetorical sections of some workplace genres in CE (e.g. Technical Memoranda, Reports). The move analysis of the two main genres in OCAW-CE is one of the main contributions of this study as the discipline-specific nature of my corpus has allowed me to characterize these assignments more precisely than has been done previously as pedagogical genres within the field of engineering.

The move analysis of the CS and MR sub-corpora is based on a corpus of a substantial number of actual student assignments from the same discipline. Discussions with subject lecturers as well as course documentation informed the move analysis process, making this analysis a much more reliable guide than other studies which did not consider contextual factors while analysing the Moves and Steps (Parkinson 2017) and based their analysis on a corpus containing multiple disciplines. It may be possible that in such corpora no one discipline has as many exemplars as may be required for genre analysis. By conducting a detailed move analysis of the Case Study and Methodology Recount genre, I have addressed RQ 1.

Many studies (e.g. Basturkmen 2009, 2012, Kanoksilapatham 2005, 2012, 2015, Martín and Pérez 2014, McGrath 2016, Samraj 2005) continue the Swalesian genre analysis tradition which has long been associated with the various sections of the Research Article genre. Although move analysis of RAs has proved useful to understand the macrostructure and linguistic features employed in professional writing for an audience of peers, more investigations of pedagogical genres are required. Student assignments and RAs are written for different purposes and for different audiences, and therefore the research into the language and structure of RAs is not always directly relevant to student assignment writers. Although over a hundred research articles on student assignments in BAWE and some work on MICUSP exist, contextual studies like the one I have undertaken are of value for the students in EFL contexts because coursework in these contexts might differ from that found elsewhere. The two genres that I examined in this study, Case Studies and Methodology Recounts, might show variations from the exemplars included in the BAWE, MICUSP or other corpora due to national, institutional or department requirements. This is a reasonable assumption given that these genres are produced in lingua franca contexts in the 'expanding circle' where the academic requirements might differ from those of L1 or ESL contexts located in the 'inner' and 'outer' circles.

Although it is unlikely that most Civil Engineering students from my institution will ever write Research Articles, they are very likely to need to produce written reports in the course of their work (Conrad 2017). Therefore some of the genres assigned to CE students simulate workplace genres in anticipation of workplace requirements. Civil Engineering Case Studies and Methodology Recounts are examples of such genres that prepare students for the type of writing they will have to do as professional engineers. Students need to be aware of the expectations of the audience and the purpose of their texts. For example, a successful Engineering Case Study, both in the workplace and as a pedagogical genre, needs to contain recommendations. In the Civil Engineering world of work, it should be noted that recommendations are key (Conrad 2018) and since the pedagogical genre is a simulation of the real genre, it is crucial that students provide recommendations in such assignments. Readers may not even read the rest of the report in great detail since the main purpose of

professional reports is to make recommendations. My specialist informants also confirmed that they awarded a lower grade to many of the CS assignments because they did not contain explicit recommendations. On being asked whether students were explicitly told about the need to make recommendations, subject teachers responded that students should 'know' that they need to suggest enhancements when conducting Case Studies (see Section 4.4). What can be concluded from this is that genre awareness and the need for explicitly pointing out the rhetorical stages of a genre are important for both faculty and students. Hardy and Friginal (2016) also emphasise the need to promote genre awareness amongst students and teaching staff. The analyses conducted in this study have helped to reveal important pedagogical points of this type.

After having described the two main genres in OCAW-CE through a move analysis, I described the phraseology and phrasal complexity features of the Case Study and Methodology Recount genres. To address the second research question on the similarities and differences in phraseology between the two genres, Lexical Bundles in the Case Study (CS) and Methodology Recount (MR) sub-corpora were retrieved and examined. The LB structures were examined using Biber et al.'s (1999) framework and they were also categorized according to their functions, using Hyland's framework (2008). LBs containing noun and preposition phrases (NP and PP) were predominant in both sub-corpora and this is in accord with research findings of earlier studies (Biber and Gray 2013, 2016, Hyland 2008) which report that academic texts tend to employ NP and PP bundles. This finding indicates that Omani assignments which met department requirements in this institution were conforming to the norms of academic writing, at least in part.

An association between the structure of LBs and the functions they fulfilled in the sub-corpora was also discerned; for example, the *(NP+) other modifier* structure was mainly used to express the location of objects in the MR assignments (e.g. *the sand in the*). The results of this analysis point to the need for additional categories, the *Noun Phrase* and *(VP+) other modifier* structure, to be added to the LB categories in Biber et al.'s (1999) framework, since they were found to be frequent enough to be categorized separately (e.g. *a smooth uniform paste, comes in contact with*). As I have already pointed out (see Section 2.4.5), researchers

have reported the frequency of phrasal complexity features in scientific writing (Biber and Gray 2016, Gardner, Nesi and Biber 2018).

In terms of function, Research Oriented bundles predominated in both sub-corpora; this propensity towards employing more RO bundles in expert and student writing in the hard sciences is also reported in the literature (Hyland 2008, Hyland 2012, Rezoug and Vincent 2018, Tribble and Wingate 2013). However, an important generic difference revealed by this analysis is that the percentage of Text-Oriented bundles in the CS assignments is double that found in the MRs. This is most likely due to the greater length of these assignments, leading to the need for more reader guidance, or because of the greater number of headings and sub-headings in the MR texts, which also help with text-orientation. CS texts were found to favour Topic, Description and Structuring bundles while MR texts favoured Procedure, Quantification and Topic bundles. These preferences are explained by the nature of the tasks students undertake in these genres. For example, it is hardly surprising that LBs realizing Procedure, Quantification and Topic bundles abound in MR assignments since these recount experimental procedures on the various topics assigned by the subject teachers and involve calculations to arrive at results.

Even within the functional sub-categories, genre variation was found; for example, Location bundles in the CS texts are used to indicate time (e.g. *the beginning of the*) while in the MR texts, they are used to show physical location of objects (e.g. *end of the block*). These types of details have value for teaching and learning since both teachers and students can be made aware of the type of Location bundle useful for Case Studies and Methodology Recounts. Participant-Oriented bundles have a negligible presence in both sub-corpora; Rezoug and Vincent's (2018) study which also investigated student writing, although not in a similar English as a lingua franca context as Oman, found similar results. The lack of stance features in student writing is possibly due to the fact that students do not need to adopt a stance like the experts because of the communicative intent of the genre they are attempting. It may therefore be less important to focus on stance features in the teaching of these genres.

The main findings of the analyses to answer the third research question will now be discussed. The similarities and differences in phrasal complexity between the CS and MR genres were investigated by examining some of the linguistic features that contribute to the density of texts. The features that were examined in these analyses are noun strings, nominalisations, attributive adjectives and the noun categories in the pattern *the N of (the) N*. Phrasal complexity features such as noun strings, nominalisations and attributive adjectives were found to be more frequent in the CS sub-corpus. The proportion of tokens in *the N of (the) N* pattern was found to be slightly higher in the MR than in the CS sub-corpus. Although phrasal complexity features were found in both sub-corpora, variations in the frequency and types of these features are apparent (see Chapter 6).

Since the analysis of LBs indicated some interesting differences in the noun categories used by the two genres, I investigated whether there was any genre influence involved in their selection. Comparison of the noun categories based on Biber's (2006) semantic noun category framework (see Section 6.3) shows that there are significant differences in the noun categories in *the N of (the) N* pattern preferred by the two types of assignments. For example, in the first position in *the N of (the) N* pattern, Abstract Nouns are preferred by the CS texts while Quantity and Concrete Nouns are preferred by the MR texts. The importance of the string, *the N of (the) N*, can be made more meaningful for students by drawing their attention to these distinct genre preferences for semantic noun categories in CS and MR assignments. A list of the most frequently used noun categories in this position may also help students internalize this structure by associating this structure with particular lexis to convey distinct meanings.

Similarly, I detected statistically significant differences in the sequencing of noun categories within *the N of (the) N* pattern. When the frequency of these noun sequences were compared, I found that CS texts use the Abstract-Abstract pattern more frequently than the MR texts while the MR texts use the Quantity-Concrete pattern more frequently. These differences in the use of noun categories are associated with the differing functions and purposes of the two genres. For example, the use of the Quantity-Concrete pattern in the MR assignments (e.g. *the mass of the sand, the volume of the given cube*) reflects the students'

need to measure and quantify objects or experiment samples in the course of doing their lab experiments. Similarly, the use of the Abstract-Abstract sequence within the pattern in the CS texts (e.g. *the definition of sustainability, the necessity of sustainable development*) reflects the students' need to define concepts or review the literature. The pedagogical value of these findings is that students can be made conscious of how these distinct sequences of semantic noun categories in *the N of (the) N* can be used to realise particular communicative functions.

A second area of investigation relating to complexity was the number of noun pre-modifiers used before the Head Noun. As shown in Table 6.1, this does not usually exceed two in either sub-corpus. Montero (1996) reports that this is the case with Computer Science writing but not many comparisons can be made to other studies because surprisingly, barring a few exceptions (Biber and Gray 2016, Gardner, Nesi and Biber 2018, Montero 1996, Parkinson 2015), not much research attention has been directed to noun strings in academic writing. These findings can directly inform the teaching material for CE students who can be made aware of the functions of strings that feature in the two genres. They can also be made aware that stringing too many nouns together may cause ambiguity.

The results of these analyses of Phrasal Complexity features and *the N of (the) N* pattern use in the two sub-corpora suggests that the two genres exhibit two types of density. Gardner, Nesi and Biber (2018) report that assignments in the Hard Sciences in the BAWE corpus tend to use Quantity and Concrete nouns and also more prenominal modifiers, while assignments in the Social Sciences use nominalisations and attributive adjectives to achieve density. My findings indicate that the MR assignments tend towards the first type of 'procedural density' while the CS assignments favour linguistic features such as nominalisations and attributive adjectives which generally characterize texts in the Social Sciences and which show the second type of density. It is worth pointing out here that Case Studies were mostly found in the business disciplines in BAWE (Nesi and Gardner 2018) and that Engineering Case Studies generally tend to be business oriented (see 2.8.4). In OCAW-CE too, the Case Studies tend not to be as technical as the Methodology Recounts as they deal with wider issues such as the environment or pollution, addressing the concerns

of those both within and outside the CE discourse community. The 'flavour' of Engineering CS writing may perhaps account for the second type of density featuring in this type of assignment.

These findings clearly indicate that a general syllabus for all genres is not practical, rather it would be more fruitful to focus on the linguistic features required for a specific genre. It clearly demonstrates that contextual factors such as the purpose and nature of the genre need to be investigated so that assumptions about such things as the 'hard science' nature of all Engineering genres can be avoided. This is especially important for students in EFL contexts such as the one I have described in this study, since teaching and learning an overwhelming array of linguistic features without pointing out their association with certain genres may not be very effective.

The main conclusion that can be reached from these findings is that genre variation is an important factor which influences the choice of linguistic devices. Some studies (e.g. Chen and Baker 2016) have compared expert genres like Research Articles to pedagogical genres without taking into consideration their different communicative purposes. They then recommend that students should blindly emulate the 'gold standard' of the Research Article genre. The results of my investigation have demonstrated that student genres differ from the RA genre, and that language features differ even between two genres within a single discipline.

7.3. Limitations

It is important to acknowledge some of the limitations of this study so that they may be addressed in future work. The first of these relates to the level of writing. Some of the students who contributed their assignments to the corpus were not very proficient writers. Hence, the assignments that subject lecturers categorized as good assignments might only be 'relatively' proficient (as they themselves acknowledged). Although from the point of view of a language specialist, these texts may be imperfect, it should be borne in mind that these texts met the requirements of the course. Since the study aims were to characterize the language needs of students and these texts embody these needs, they have served their

purpose well. Seen from this perspective, the 'imperfection' of the assignments in OCAW-CE might not be drawback at all, but just a reflection of real world Engineering writing in EFL contexts. Expert writing in Engineering might also contain some of these imperfections while remaining acceptable to the discourse community in terms of the community's priorities.

A further limitation of this study may be that all the assignments in OCAW-CE have been collected from a single institution. Institutional and departmental cultures may have influenced the way these assignments are organized and the choice of lexico-grammatical features. For example, students in this institution are often advised not to use the first person and are encouraged to use imperatives while recounting experiments. This may not be the case in other contexts when the same types of texts could be written using a variety of features such as passive voice or modals in Methods sections. Readers therefore need to interpret these findings in the light of their own contexts.

There are also some drawbacks associated with the methods used to investigate phraseology, principally the automatic methodology used to retrieve Lexical Bundles, which did not allow for variation in the ordering of words within the frequent four-word strings.

7.4. Future Directions

This section will discuss the possibilities of future research that can build on the study reported in this thesis. Through this study, the possibility of compiling corpora 'from scratch' to inform the learning resources and teaching approaches suited to particular contexts has been demonstrated. The processes involved in designing and compiling the OCAW-CE have been detailed not only in this thesis but also in a book chapter (Mathew, Nesi and Vincent 2019). This information and the findings of this study may possibly encourage more EAP instructors, researchers and course book developers to compile specialized corpora as a means of investigating the language items students need to use. Since studies such as (Gardner, Nesi and Biber (2018) and Staples et al. (2016) continue to demonstrate that disciplinary and genre variation are factors that need to be considered by all stakeholders

involved in the learning and teaching of academic writing, the importance of first identifying the needs of students by compiling relevant corpora becomes apparent.

The OCAW-CE corpus contains assignments in the Civil Engineering discipline collected from a single institution. The very name of the corpus, the Omani Corpus of Academic Writing, implies that this project has only just begun, and should be extended to represent assignments from a wider range of disciplines. Assignments might also be collected from other institutions within Oman such as Sultan Qaboos University (SQU) to enhance the representativeness of the corpus. This is a realistic next step; since the Writing Centre at MEC collaborates with the Centre at SQU, a jointly submitted funded project sponsored by The Research Council of Oman could be the next stage towards this enhancement of OCAW-CE.

A further possibility for developing the project would be to include corpora already in existence, such as the ACE corpus (Rezoug and Vincent 2018) with the aim of creating a student corpus for 'English as a lingua franca' countries in the 'expanding circle' (ELF Student Corpus). Since there have been very few attempts to identify the genres assigned to students in universities in EFL contexts, this project could be the initiation of much-needed research to help struggling students cope with their disciplinary writing requirements. If this corpus is publicly archived, researchers not only based in the Arab-speaking world but from the 'expanding circles' generally could come together to create a wave of research interest in disciplinary writing.

The outcomes of such research can be used by EAP instructors, students and course book developers to avoid 'shots in the dark', a hallmark of general English textbooks purporting to support students with their written assignments. Although many course books today are corpus-based or corpus-informed (e.g. Garnet Education textbooks), they could better suit the needs of students if they are based on corpora containing texts written by the kind of students they aim to support. For example, a course book developer designing an ESP syllabus for CE students might customize the book for the Middle East based on the OCAW-CE corpus or its successors. Basing the design of ESP courses on context-specific

corpora may perhaps narrow the gap between the syllabus and learning materials on the one hand and what the students actually need on the other.

It is true that compiling corpora consisting of 'occluded' (Swales 1996: 46) genres such as student assignments may be challenging, but such projects need to be started for the benefit of millions of students worldwide.

7.5. Pedagogical Implications

The linguistic descriptions of the two main genres in Civil Engineering undertaken in this study serve the main purpose of identifying potential student needs so that an appropriate ESP for Civil Engineering syllabus can be designed based on key linguistic and organizational features that need to be taught. This study will not 'fade away before offering well-articulated pedagogical applications' (Swales 2019: 78); in this section I will briefly describe some of the pedagogical applications of my findings some of which can also be adapted for students in other disciplines. Since this study was based on a corpus of student writing produced in Oman, one of the countries in the 'expanding circle' in which English is the lingua franca, these recommendations may be especially suitable for students in similar contexts to those of the research site.

7.5.1. Genre Pedagogy

A few studies (Parkinson 2017, Tribble and Wingate 2013) have offered explicit teaching applications for students based on genres of university student writing. This section will outline some pedagogical applications for the move analysis of the CS and MR genres described in Chapter 4.

1. Consciousness-raising about the genres that students are expected to attempt is one way of helping students realize the rhetorical structure of genres and the communicative purpose behind each of the sections. The lack of understanding of these unfamiliar genres is 'the most pressing problem students face in a university system' (Tribble and Wingate 2013: 309). To help remedy this problem, students can be provided with texts and asked to identify the moves and steps in sample assignments (Devitt 2015, Tribble and Wingate 2013). However,

care needs to be taken to alert students to the dangers of mistaking the move analysis for a rigid unvarying checklist. As Johns (2011: 64) suggests, 'students should view texts as both temporarily structured and evolving…they should draw from prior knowledge of texts but be open to the demands of a new situation or assignment.'

2. Providing some hints and support for the lexico-grammatical features involved in realizing the Moves and Steps can help students who may not have the syntactic and lexical resources to accomplish this task. For example, my experience in the Writing Centre has shown me that many students do not use Location Statements (Swales and Feak 2012) to describe a figure or table included in their assignment. Students can be provided with the lexis and made familiar with the information they usually need to include in this kind of statement (e.g. *Table/Figure # provides/illustrates/shows + summary of the data provided*).

Genre-pedagogy is an approach that suits students of both higher and lower proficiency levels. For L2 learners struggling to cope with basic academic writing conventions, it offers the dual benefit of allowing the student to understand genre conventions as well as to focus on relevant linguistic features. Lower proficiency students can be alerted to the lexico-grammatical features associated with Moves and Steps. This can reduce the burden of learning a wider range of grammar and lexis. The lexico-grammatical features thus identified can inform the syllabus of the Foundation Programme where remediation of basic grammar and mechanics of writing is undertaken (see Section 1.1).

Higher proficiency students can also benefit from genre knowledge and associated lexico-grammatical features, even though they may possess a wider repertoire of linguistic devices with which they can realise the communicative purposes of each Move and Step. As discussed in Section 1.1, there may be room for improvement even in the work of such students; as lecturers pointed out, some of the steps which they considered mandatory (see Section 4.2), were not included by the students whose work is represented in this corpus. These students can therefore benefit from explicit discussion of genre requirements. The Centre for Academic Writing can disseminate this information to academic departments and

prepare learning materials that are tailored to help students meet their specific disciplinary writing needs.

3. Training the subject lecturers to identify the Moves and Steps in assignments is another way of encouraging them to commit to helping students with the structuring and writing of their assignments (Hardy and Friginal 2016) and also to become aware of their assignment design. This approach can be useful since Writing Centres (such as the one at the research site) usually do not have many faculty members and the number of students may vastly outweigh the help that can be provided to them. At the research site, meetings have already been held with academic departments and the training of subject lecturers has been agreed in principle. The subject lecturers acknowledged that conducting a move analysis for part genres such as the Methodology sections in project reports would be very useful indeed, since the instructions they gave were often inconsistent and this affected the quality of student work. For example, they explained that students are expected to justify their methodology by referring to the literature but are not given explicit instructions on how to do so; subject lecturers assume genre knowledge on the part of the students and often find it difficult to explicate their own tacit understanding of the genre. In addition to training subject teachers on the explicit communication of the staging, and the Moves and Steps of genres, another useful intervention could be to include these instructions in the assignment briefs. This approach would need to be systematically implemented by trained subject lecturers, and monitored periodically by language teachers.

7.5.2. Teaching applications for Lexical Bundles

My investigation revealed that there is genre variation in the LBs preferred by the two types of assignments. Before the pedagogical applications can be thought about, however, the LBs worth teaching need to be identified. Although they may be a very frequent feature in student writing, LBs are often structurally incomplete and many of them make little sense when isolated from their context. For example, the LB ‘of cement parts of’ is not a complete structural unit and would not be easy for learners to remember.

Therefore, decisions regarding the usefulness of the LBs chosen for teaching purposes must be well thought out. Simpson Vlach and Ellis (2010) asked EAP instructors to choose the LBs worth teaching in order to decide which ones are most useful for students. The LBs generated from the CS and MR texts can be shown to language tutors as well as subject teachers who can decide which are the most relevant . For example, Writing Centre lecturers can identify that LBs containing directives are useful for students (e.g. *weigh and record the, determine the mass of, increase the water content*) in the recount of experiments.

Another option would be to identify the sub-functions of LBs (Section 5.3.1) and ask students to notice a frequently repeated pattern which the LB is an instantiation of. The list of words belonging to a semantic category that always fits into the variable slots within this pattern can also be pointed out. For example, LBs used in MR assignments to describe measurement or quantification (*the weight of the, the mass of the*) or to indicate location of an object (e.g. *the base of the, the centre of the*) have specific functions, associated structures and a restricted range of lexical items or semantic categories (underlined in the example that follows) that are used in the variable slots (e.g. *the (mass/weight/volume) of the*). This is another area where there is potential for consciousness raising about syntactic structure and vocabulary. Edited concordance lines can help students to notice the semantic category and POS of words that come after this type of LB. For example, Table 7.1 from the MR sub-corpus shows lines containing concrete nouns (*cylinder, soil, beaker, hole, container*) that invariably follow LBs with the pattern *the (mass/weight/volume) of the*.

falling close the shutter 5. Determine	**the mass of the**	calibrating cylinder filled with sand and
The volume of the hole is equal to	**the mass of the**	sand filled in the hole divided by its dry
nozzle of Eureka Can. Now note down	**the weight of the**	beaker with the flowed water as 'W2 =
a small cylindrical pit is excavated and	**the weight of the**	soil excavated from the pit is measured.
of the excavated soil is determined.	**The volume of the**	hole is determined by filling it with clean,
container together. 4.4.3 Calculate	**the volume of the**	container. Weigh the container. 4.4.4

Table 7.1. LBs with the pattern *the N of the* followed by concrete nouns

The pedagogical implications of my findings are that students can be introduced to the LBs belonging to the genres that they need imminent help with, within the constraints of teachability that I have just discussed. By providing some examples of how at least some of these LBs may have pedagogical applications, I have perhaps avoided the dangers of making

this one of the 'ringfenced studies of lexical bundles' that Swales complains about (2019: 76) which merely generate lists of LBs without any hint of teaching applications.

7.5.3. Teaching Linguistic Features that contribute to density

In Chapter 6, I examined some linguistic devices that contribute to the density of texts, such as Phrasal Complexity features and *the N of (the) N* pattern. Since these features have been identified as being present in these genres to varying degrees, some of the more frequent ones can be chosen for explicit teaching or consciousness-raising activities. For example, students can be shown the noun categories used as pre-modifiers in noun strings generated from OCAW-CE identified in Section 6.2.2.1, and encouraged to notice how tangible or concrete nouns (e.g. *wood, sand*), intangible or abstract nouns (e.g. *risk, quality*) and nominalisations (e.g. *distribution, construction*) are used as pre-modifiers. The way they make academic writing denser by removing prepositional markers and restrictive relative clauses could also be useful information for students.

Other features that merit explicit teaching include nominalisations and shell nouns, used to achieve lexical density and increase textual cohesion. In fact, Omani students often fail to use nominalisations because of lack of awareness, as in Example 1, where *determine* is used in a sub-heading instead of the nominalized version *determination*.

(1) Determine of the bulk density of soil (Sand Replacement Method)

Aktas and Cortes (2008) advocate the explicit teaching of shell nouns, using illustrative examples. They point out that shell nouns are used anaphorically and cataphorically by proficient writers to draft succinct text, and to carry forward an argument and explanation. Some examples of shell noun use are evident in the corpus (see Section 2.8.6) and students will surely benefit by knowing how to use this feature. In my experience in the Writing Centre, I have come across many student texts which lack cohesion; often, when students do try to bring some unity to their texts, they are unaware of methods to improve the cohesion and consequently overuse signal phrases such as 'moreover', 'however' and 'in addition' to

the point that they occur in every sentence. Explicit teaching of other methods to enhance cohesion such as the use of shell nouns could be a solution.

Similarly, *the N of (the) N* pattern has proven to be a frequent pattern in both genres, comprising about 5 to 6% of the total tokens in both sub-corpora. This has considerable pedagogical implications since students' attention can be drawn to common patterns in the MR genre (e.g. Quantity + Concrete; *the mass of the calibrating cylinder, the weight of the beaker*) and the CS genre (e.g. Abstract + Abstract; *the quality of the air, the objective of the project*). The semantic categories and types of nouns used in these sequences within *the N of (the) N* pattern in the two genres could also be pointed out. For example, a list of the most frequently used Quantity nouns in the first position and Concrete nouns in the second position in MR assignments can be provided to students (see Table 6.5) and then they could be shown concordance lines with this sequence so that they could also view the expanded context.

7.5.3.1. Data-driven Learning

Linguistic features that have been identified as important for learners can be taught inductively and deductively via DDL (Data Driven Learning) and other methods. Data Driven Learning (DDL) is a pedagogical tool developed by Tim Johns in the 1980s which involves students directly interacting with corpora. In the 'hard' form of DDL, more autonomy is given to students and they are encouraged to explore corpora on their own, while the 'soft' version involves creating carefully chosen concordance lines to help students with particular language features they have difficulty in using correctly or appropriately (Gabrielatos, 2005; Vincent and Nesi 2018). The improvement in student writing as a result of using this type of corpus-based exercise (soft and hard DDL) is reported by researchers (Boulton and Cobb 2017, Friginal 2013).

DDL could be an appropriate approach to familiarize students with some of the language features and frequent phrases that students need to be familiar with while attempting the genres described in this study. This pedagogical approach can also be used to help students with some problems which are 'untreatable' (Ferris 2002), that is, those that learners cannot find in a reference source (Vincent and Nesi 2018: 5). A case in point would

be the use of shell nouns. DDL can also resolve some of the problems associated with familiarizing students with phrases and 'teachable' or 'must know' LBs out of their context. The 'soft' DDL approach is particularly suitable for students at MEC because many of them are not advanced learners of English and exposing them to relevant and limited selections of the item to be taught has a better chance of succeeding than exposing them to overwhelming amounts of authentic data.

Using DDL, students can be trained to notice relevant patterns in the words used before and after frequently repeated strings. For example, in CS texts, three LB strings with the pattern *length + of* were retrieved. A search of the sub-corpus revealed that this string has a normalized frequency of about 485 pmw and occurs in about 40% of the texts. Table 7.2 provides a list of selected concordance lines that can be used to encourage students to discern *the N of (the) N* pattern in which this string is found. Since many students have a problem with the use of definite articles, they can be encouraged to work out the grammatical rules underlying the definite article or examine the adjectives and other intervening words (e.g. *total*) used before *length of* or the Concrete (e.g. *truck, fins*) nouns following it.

Kibbitzers (Johns 1997), belonging to the category of 'soft' DDL, are another way of resolving particular problems faced by students. These can be created by EAP instructors to address specific problems encountered by students in the use of the linguistic items required by them. Students can also be directed to other sources such as the BAWE Quick Links or Tim Johns' kibbitzers where there are ready-made concordances available for some language features. Boulton (2017: 483) speaks about how DDL still remains a 'marginal practice' but with findings that identify the precise needs of students and with resources such as the ready-made DDL materials designed by Vincent and Nesi (2018), DDL might well become a 'mainstream practice'. Since the staff members in the Writing Centre are not corpus-literate, this ready-made solution is perhaps the most viable option, and after some training it can also be made available to subject teachers.

the soil type within the project and the	**length of**	the truck from 8 meters to 20 meters and the
average delay per vehicle is with a queue	**length of**	3 vehicles. Thus, the average delay per
King Abdullah Financial District and the	**length of**	this line is approximately 29.6 km, and
Arabia (Paul Snowden, 2014). The total	**length of**	the track will be (279-km) and is scheduled
City and urban track The calculable total	**length of**	the Oman National railway network is (2244-
(km), passenger and freight The total	**length of**	Doha metro is (234-km) network above and
five of A380 aircraft and the total	**length of**	stainless steel bull nose fins is 293 times the

Table 7. 2. Concordance lines containing the phrase *length of* within *the N of (the) N* pattern

7.6. Significance of the study

I end this thesis by summarizing the value of the research I have undertaken. By creating a corpus of student writing, I have demonstrated that with a minimal amount of training it is possible for EAP instructors and researchers to do the same, albeit on a smaller scale because of practical limitations such as lack of time and resources. If more corpora of this type are compiled, they can fuel the next wave of research in pedagogic genres, since corpus research into discipline-specific student writing is woefully lacking, especially in the 'expanding circle' group of countries that have adopted English as their lingua franca. The need for these types of investigations into student writing should be measured in terms of the millions of students already enrolled in university and their ever-increasing numbers each year in institutions like MEC scattered over the world.

In this thesis, I have argued that Oman can be grouped with other countries in the 'expanding circle' which have a unique identity in terms of the academic writing support required by students for successful communication. The workplaces in these countries have adopted English as the means of communication, and their higher education institutions, which are mainly affiliates of Western universities located in L1 countries, have adopted English as the medium of instruction. The adoption of English as the lingua franca in the world of work and business in these countries has been prompted by two main factors: the presence of a large expatriate workforce and participation in international trade. However, the language of instruction in schools is the local language of these countries, and the first language of most of the expatriate workforce is not English. This situation leads to a unique university environment where most students need to adapt to the academic writing

conventions of a Western university without having the benefit of a Western model of school education which could have possibly equipped them with the language proficiency required for engaging in disciplinary discourse in English. Many students find it difficult to transition from an Arabic-speaking environment to a Western university culture with its expectations of linguistic competence. It is in this unique context that I have placed Oman and the research site of this study.

I have suggested that a practical approach to supporting students in these contexts would be to narrow the range of linguistic resources required in order to communicate effectively in their disciplinary genres. The findings reported in Chapters 4, 5 and 6 indicate that proficient writers in these contexts use a narrower range of linguistic items than L1 English students, but have nevertheless been successful in producing assignments that have met departmental requirements. Research into the genres that students produce in specific institutions and disciplines can result in in-depth descriptions, informed not only by coursework prompts but by student assignments, course documentation and discussions with specialist informants (see 2.3.4). The study reported in this thesis fits this category of research studies by describing the primary genres and their phraseology in Civil Engineering in a single institution. The findings of this study may possibly be applicable to similar contexts in the 'expanding circle' where English is the lingua franca.

The analyses which I have undertaken to provide these genre descriptions have considerable pedagogical implications, and can claim to have advanced our knowledge in little-explored but critical research areas.

References

Ädel, A. and Erman, B. (2012) 'Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach'. *English for Specific Purposes* 31 (2), 81-92

Aktas, R. N. and Cortes, V. (2008) 'Shell nouns as cohesive devices in published and ESL student writing'. *Journal of English for Academic Purposes* 7, 3-14

Alshahrani, S. (2015) 'A cross-linguistic analysis of interactive metadiscourse devices employment in native English and Arab ESL academic writings'. *Theory and Practice in Language Studies* 5 (8), 1535-1542

Al Jadidi, H. and Sangunietti, J. (2010) 'Characteristic pedagogical styles of bilingual and monolingual English teachers'. *International Journal of Arts and Sciences,* 3 (16), 131-147

Al Jardhani, S. S. S. (2017) 'English language curriculum evaluation in Oman'. *International Journal of English Linguistics* 2 (5), 40-44

Al Mahrooqi, R. and Denman, C.J. (2017) 'Investigating FL attrition among Omani teachers of English: Implications for educational reform in Oman and the Arab Gulf'. in *Challenges to Education in the GCC during the 21st Century* ed. by Mahboob, A. and Elyas, T. Jeddah, KSA: Gulf Research Centre Cambridge, 19-46

Alsop, S. and Nesi, H. (2009) 'Issues in the development of the British Academic Written English (BAWE) corpus'. *Corpora* 4 (1), 71-83

Altenberg, B. (1998) 'On the phraseology of spoken English: the evidence of recurrent word-combinations'. in *A Phraseology: Theory Analysis and Applications.* ed. by Cowie, P. Oxford: Oxford University Press, 101-122

Anthony, L. (1999) 'Writing research article introductions in software engineering: How accurate is a standard model?' *Professional Communication, IEEE Transactions* 42 (1), 38-46

Artemeva, N. and Freedman, A. (2008) *Rhetorical Genre Studies and Beyond*. Winnipeg, MB: Inkshed Publications

Artemeva, N. (2009) 'Stories of becoming: A study of novice engineers learning genres of their profession'. in *Genre in a Changing world. Perspectives on Writing*. ed. by Bazerman, C., Bonini, A. and Figueiredo, D. Fort Collins, Colorado: The WAC Clearinghouse and Parlor Press, 158-178

Artstein, R. and Poesio, M. (2008) 'Inter-coder agreement for computational linguistics'. *Computational Linguistics* 34 (4), 555-596

Banks, D. (2005) 'On the historical origins of nominalized process in scientific text'. *English for Specific Purposes* 24, 347–357

Bartolic, L. (1978) ‘Nominal compounds in technical English’. in *English for Specific Purposes, Science and Technology*. ed. by Trimble, T. and Drobnic, T. English Language Institute: Oregon State University, 257-277

Basturkmen, H. (2012) ‘A genre-based investigation of discussion sections of research articles in dentistry and disciplinary variation’. *English for Academic Purposes* 11(2), 134-144

Basturkmen, H. (2009) ‘Commenting on results in published research articles and masters dissertations in language teaching’. *Journal of English for Academic Purposes* 8 (4), 241-251

Bazerman, C. (1981) ‘What written knowledge does: Three examples of academic discourse’. *Phil. Soc. Sci.* 11, 361-387

Becher, T. (1989) *Academic Tribes and Territories*. Milton Keynes: The Society for Research into Higher Education and Open University Press

Becher, T. (1994) ‘The significance of disciplinary differences’. *Studies in Higher Education* 19 (2), 151-161

Becher, T. and Trowler, P. (2001) *Academic Tribes and Territories: Intellectual Enquiry and the Cultures of Disciplines*. Buckingham: Open University Press/SRHE

Bernhardt, S. A. (1985) ‘Writing across the curriculum at one university: A survey of faculty members and students’. *AbE Bulletin* 82, 55-59

Bestgen, Y. (2019) ‘Comparing lexical bundles across corpora of different sizes: The Zipfian problem’. *Journal of Quantitative Linguistics* 1-19

Biber, D. (1988) *Variation across Speech and Writing.* Cambridge: Cambridge University Press

Biber, D. and Barbieri, F. (2007) ‘Lexical bundles in university spoken and written register’. *English for Specific Purposes* 26 (3), 263-286

Biber, D. (2006) *University Language: A Corpus-based Study of Spoken and Written registers*. Amsterdam: John Benjamins

Biber, D., and Gray, B. (2013) ‘Nominalizing the verb phrase in academic science writing’. in *The English Verb Phrase: Corpus Methodology and Current Change*. ed. by Aarts, B. and Leech, G. Cambridge, UK: Cambridge University, 99-132

Biber, D. and Gray, B. (2016) *Grammatical Complexity in Academic English: Linguistic Change in Writing*. Cambridge: Cambridge University Press

Biber, D., Conrad, S., and Cortes, V. (2004). ‘If you look at…: Lexical bundles in university teaching and textbooks’. *Applied linguistics* 25 (3), 371-405

Biber, D., Conrad, S., Reppen, R., Byrd, P., Helt, M., Clark, V., Cortes, V., Csomay, E., & Urzua, A. (2004) *Representing Language Use in the University: Analysis of the TOEFL 2000*

Spoken and Written Academic Language Corpus. TOEFL Monograph Series Princeton, New Jersey: Educational Testing Service

Biber, D. and Conrad, S. (2009) *Genre, Register, Style*. Cambridge: CUP

Biber, D., Gray, B., and Poonpon, K. (2011) 'Should we use characteristics of conversation to measure grammatical complexity in L2 writing development?' *TESOL Quarterly* 45, 5-35

Biber, D., Johansson, S., Leech, G., Conrad, S., and Finnegan, E. (1999) *The Longman Grammar of Spoken and Written English*. London: Longman

Biber, D. and Zhang, M. (2018) 'Expressing evaluation without grammatical stance: informational persuasion on the web'. *Corpora* 13 (1), 97-123

Billig, M. (2013) *Learn to Write Badly: How to Succeed in the Social Sciences*. Cambridge, UK: Cambridge University Press

Blaikie, N. (2003) *Analyzing Quantitative Data: From Description to Explanation*. Wiltshire, UK: Sage

Bloor, T. and Bloor, M. (2013) *The Functional Analysis of English*. London: Routledge

Bolinger, D. (1976) 'Meaning and memory'. *Forum Linguisticum*, 1, 1–14

Boulton, A. (2017) 'Research timeline: Corpora in language teaching and learning'. *Language Teaching* 50, 483-506

Braxton, J. (1995) 'Disciplines with an affinity for the improvement of undergraduate education'. in *Disciplinary Differences in Teaching and Learning*. ed. by Hativa, N. and Marincovich, M. San Francisco: Jossey-Bass Publishers, 59-64

Bridgeman, B. and Carlson, S. (1984) 'Survey of academic writing tasks'. *Written Communication* 1, 247-280

Cai, J. (2016) 'An exploratory study on an integrated genre-based approach for the instruction of academic lexical phrases'. *Journal of English for Academic Purposes* 24, 58-74

Canseco, G. and Byrd, P. (1989) 'Writing required in graduate courses in business administration'. *TESOL Quarterly* 23, 305–316

Carter, M. (2007) 'Ways of knowing, doing and writing in the disciplines'. *College Composition and Communication* 58, 385-418

Casanave, C. and Hubbard, P. (1992) 'The writing assignments and writing problems of doctoral students: Faculty perceptions, pedagogical issues, and needed research'. *English for Specific Purposes* 11, 33–49

Charles, M. (2003) 'This mystery…': a corpus-based study of the use of nouns to construct stance in theses from two contrasting disciplines'. *Journal of English for Academic Purposes* 2, 313–326

Chen, Y.H. and Baker, P. (2010) 'Lexical bundles in L1 and L2 academic writing'. *Language Learning and Technology* 14 (2), 30-49

Chen, Y.H. and Baker, P. (2016) 'Investigating criterial discourse features across second language development: Lexical bundles in rated learner essays, CEFR B1, B2 and C1'. *Applied Linguistics* 37(6), 849-880

Cheng, A. (2007) 'Transferring generic features and re-contextualizing genre awareness: Understanding writing performance in the ESP genre-based literacy framework'. *English for Specific Purposes* 26, 287–307

Conrad, S. (2018) 'The use of passives and impersonal style in Civil Engineering writing'. *Journal of Business and Technical Communication* 32 (1), 38-76

Conrad, S. (2017) 'A comparison of practitioner and student writing in civil engineering'. *Journal of Engineering Education* 106 (2), 191-217

Conrad, S., Kitch, W., Smith, T.R., Lamb, K., and Pfeiffer, T. (2016) Faculty practitioner collaboration for improving civil engineering students' writing skills. in *ASEE Annual Conference & Exposition.* held 26-29 June at New Orleans Convention Centre, LA, USA. [online] available from <https://www.asee.org/conferences-and-events/conferences/annual-conference/past-conferences/2016> [16 July 2019]

Conrad, S. (2014) 'Expanding multi-dimensional analysis with qualitative research techniques'. *Multi-dimensional Analysis* 25, 273-95

Cortes, V. (2002) 'Lexical bundles in freshman composition' in *Using corpora to explore linguistic variation* ed. by Reppen, R., Fitzmaurice, S.M., and Biber, D. Amsterdam: John Benjamins Publishing, 131-146

Conrad, S. and Biber, D. (2004) 'The frequency and use of lexical bundles in conversation and academic prose'. *Lexicographica* 20, 56-71

Cooper, A. and Bikowski, D. (2009) 'Writing at the graduate level: what tasks do professors actually require?' *Journal of English for Academic Purposes* 6 (3), 206-221

Cortes, V. (2013) 'The purpose of this study is to: Connecting lexical bundles and moves in research article introductions'. *Journal of English for Academic Purposes* 12 (1), 33-43

Cortes, V. (2004) 'Lexical bundles in published and student disciplinary writing: examples from history and biology'. *English for Specific Purposes* 23 (4), 397-423

Cowan, J. R. (1974) 'Lexical and syntactic research for the design of EFL reading materials'. *TESOL Quarterly* 8, 389–400

Coxhead, A. (2000) ‘A new academic word list’. *TESOL Quarterly* 34 (2), 213-238

Crompton, P. (2011) ‘Article errors in the English writing of advanced L1 Arabic learners: The role of transfer’. *Asian EFL Journal* 50 (1), 4-35

Cunningham, K.J. (2017) ‘A phraseological exploration of recent mathematics research articles through key phrase frames’. *Journal of English for Academic Purposes* 25, 71-83

De Cock, S. (1998) ‘A recurrent word combination approach to the study of formulae in the speech of native and non-native speakers of English’. *International Journal of Corpus Linguistics* 3 (1), 59–80

Denzin, N. K. (1978) *The Research Act: A Theoretical Introduction to Sociological Methods*. New York: McGraw-Hill

Devitt, A. (2004) *Writing Genres*. Carbondale: Southern Illinois University Press

Devitt, A.J. (2015) ‘Genre performances: John Swales' genre analysis and rhetorical-linguistic genre studies’. *Journal of English for Academic Purposes* 19, 44-51

Durrant, P. (2014) ‘Discipline and level specificity in university students’ written vocabulary’. *Applied Linguistics* 35 (3), 328-356

Durrant, P. (2017) ‘Lexical bundles and disciplinary variation in university students’ writing: Mapping the territories’. *Applied Linguistics* 38 (2), 165-193.

Ebeling, S.O. and Heuboeck, A. (2007) ‘Encoding document information in a corpus of student writing: The British Academic Written English corpus’. *Corpora* 2 (2), 241-256

Ebeling, S. O. and Hasselgård, H. (2015) ‘Learners' and native speakers' use of recurrent word-combinations across disciplines’. *Bergen Language and Linguistics Studies* 6, 87-106

Eblen, C. (1983) ‘Writing across-the-curriculum: a survey of a university faculty’s views and classroom practices’. *Research in the Teaching of English* 17, 343–348

Eeg-Olofson, M. and Altenberg, B. (1994) ‘Discontinuous recurrent word combinations in the London-Lund Corpus’. In Fries, U., Tottie, G. and Schneider, P. (eds.), *Proceedings of the fourteenth international conference on English language research on computerized corpora* ‘Creating and using English language corpora’. held 1993 at Zürich. Amsterdam: Rodopi, 63-77

Elander, J., Harrington, K., Norton, L., Robinson, H., and Reddy, P. (2006) ‘Complex skills and academic writing: a review of evidence about the types of learning required to meet core assessment criteria’. *Assessment & Evaluation in Higher Education* 31(1), 71-90

El Malik, A. T. and Nesi, H (2008) ‘Publishing research in a second language: The case of Sudanese contributors to international medical journals’. *Journal of English for Academic Purposes* 7 (2), 87-96

Ellis, N. C., Simpson-Vlach, R., and Maynard, C. (2008) 'Formulaic language in native and second language speakers: Psycholinguistics, Corpus Linguistics, and TESOL'. *TESOL Quarterly* 42 (3), 375–396

Ferris, D. (2002) *Treatment of Error in Second Language Student Writing*. Ann Arbor, MI: The University of Michigan Press

Firth, J. R. (1957) *Papers in Linguistics*. London: Oxford University Press

Firth, A. (1996) 'The discursive accomplishment of normality: On 'lingua franca' English and conversation analysis'. *Journal of Pragmatics* 26 (2), 237-259

Fletcher, W.H. (2012) *kfNgram. Information & Help*. [online] available from <https://www.kwicfinder.com/kfNgram/kfNgramHelp.html> [1 July 2019]

Flowerdew, J. (1993) 'Content-based language instruction in a tertiary setting'. *English for Specific Purposes* 12 (2), 121-138

Flowerdew, J. (2003) 'Signalling nouns in discourse'. *English for Specific Purposes* 22, 329-346

Flowerdew, J. and Forest, R.W. (2015) *Signalling Nouns in Academic English*. London: Cambridge University Press

Flowerdew, J. (2016) 'English for Specific Academic Purposes (ESAP) writing: Making the case'. *Writing & Pedagogy* 8 (1), 5-32

Fox, J. and Tigchelaar, M. (2015) 'Creating an engineering academic formulas list'. *Journal of Teaching English for Specific and Academic Purposes* 3 (2), 295-304

Francis, G. (1994) 'Labelling discourse: an aspect of nominal-group lexical cohesion'. in *Advances in Written Text Analysis.* ed. by Coulthard, M. New York: Routledge, 97-115

Fuoli, M. and Hommerberg, C. (2015) 'Optimising transparency, reliability and replicability: Annotation principles and inter-coder agreement in the quantification of evaluative expressions'. *Corpora* 10 (3), 315-349

Gardner, S. (2008) 'Integrating ethnographic, multidimensional, corpus linguistic and systemic functional approaches to genre description: An illustration through university history and engineering assignments' in *ESFLCW 2007: Data and Interpretation in Linguistic Analysis* ed. by Steiner, E.H. and Neumann, S. Saarbrücken: Universität des Saarlandes, 1-34

Gardner, S. (2012) 'Perspectives on the disciplinary discourses of academic argument' in *Corpus Linguistics Conference 2011*. held 20-22 July at the University of Birmingham. [online] available from <https://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2011/Paper-71.pdf> [16 July 2019]

Gardner, S., Nesi, H., and Biber, D. (2018) 'Discipline, level, genre: Integrating situational perspectives in a new MD analysis of university student writing'. *Applied Linguistics* 10, 1-30

Gebhard, M., Chen, I., and Britton, L. (2014) 'Miss, nominalization is a nominalization': English language learners' use of SFL metalanguage and their literacy practices'. *Linguistics and Education* 26, 106-125

Gillettt, A. and Hammond, A. (2009) 'Mapping the maze of assessment: An investigation into practice'. *Active Learning in Higher Education* 10, 120-137

Giltrow, J. (1994) 'Genre and the pragmatic concept of background knowledge'. in *Genre and the New Rhetoric* ed. by Freedman, A. and Medway, P. London: Taylor & Francis, 130-151

Gilquin, G., Granger, S., and Paquot, M. (2007) 'Learner corpora: The missing link in EAP pedagogy'. *Journal of English for Academic Purposes* 6 (4), 319-335

Granger, S., Dagneaux, E., Meunier, F., and Paquot, M. (2009) *ICLE: International corpus of learner English* Louvain-la-Neuve, Belgium: Presses Universitaires de Louvain

Gray, B.E. (2011) *Exploring Academic Writing through Corpus Linguistics: When Discipline Tells Only Part of the Story*. Unpublished Dissertation. Arizona: Northern Arizona University

Greenland, S., Senn, S.J., Rothman, K.J., Carlin, J.B., Poole, C., Goodman, S.N., and Altman, D.G. (2016) 'Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations'. *European Journal of Epidemiology* 31(4), 337-350

Hale, G., Taylor, C., Bridgeman, B., Carson, J., Kroll, B., and Kantor, R. (1995) 'A study of writing tasks assigned in academic degree programs'. *ETS Research Report Series* 2, i-61

Halliday, M. A. K. and Martin, J. R. (1993) *Writing Science, Literacy and Discursive Power*. London: Falmer Press

Halliday, M. A. K. and Matthiessen, C. M. I. M. (1999) *Construing Experience through Meaning: A Language-based Approach to Cognition.* London: Continuum

Halliday , M. A. K. and Matthiessen, C. M. I. M. (2004) *An Introduction to Functional Grammar*. London: Arnold

Hammersley, M. (1992) *What's Wrong with Ethnography.* London: Routledge

Hammersley, M. (2006) 'Ethnography: problems and prospects'. *Ethnography and Education* 1 (1), 3–14

Hammersley, M. and Atkinson, P. (2007) *Ethnography: Principles in Practice*. New York: Routledge

Hardy, J.A. and Romer, U. (2013) 'Revealing disciplinary variation in student writing: a multi-dimensional analysis of the Michigan Corpus of Upper-level Student Papers (MICUSP)'. *Corpora* 8 (2), 183–207

Hardy, J.A. and Friginal, E. (2016) 'Genre variation in student writing: A multi-dimensional analysis'. *Journal of English for Academic Purposes* 22, 119-131

Herzberg, B. (1986) 'The politics of discourse communities'. in *Conference on College Composition and Communication*. held 13-15 March at New Orleans. [online] available from <https://cccc.ncte.org/cccc/review/pastconventions> [16 July 2019]

Hinkel, E. (2004) *Teaching Academic ESL Writing: Practical Techniques in Vocabulary and Grammar*. NJ: Lawrence Erlbaum Associates

Horowitz, D.M. (1986a) 'Essay examination prompts and the teaching of academic writing'. *English for Specific Purposes* 5, 197-120

Horowitz, D.M. (1986b) 'What professors actually require of students: Academic tasks for the ESL classroom'. *TESOL Quarterly* 20, 445–462

House, J. (2012) 'Teaching oral skills in English as a lingua franca'. in *Principles and Practices for Teaching English As An International Language.* ed. by Alsagoff, L., Sandra Lee Mckay, S.L., Hu, G. and Renandya, W.A. New York: Routledge, 192-211

Hsu, W. (2014) 'Measuring the vocabulary load of engineering textbooks for EFL undergraduates'. *English for Specific Purposes* 33, 54-65

Huddleston, R. and Pullum, G. (2002) *The Cambridge Grammar of the English Language*. Cambridge, UK: Cambridge University Press

Hunston, S. and Francis, G. (2000) *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins Publishing

Hyland, K. (1999) 'Academic attribution: citation and the construction of disciplinary knowledge'. *Applied Linguistics* 20, 341–367

Hyland, K. (2002) *Disciplinary Discourses: Social Interactions in Academic Writing*. Michigan: University of Michigan Press

Hyland, K. (2008) 'As can be seen: Lexical bundles and disciplinary variation'. *English for Specific Purposes* 27 (1), 4–21

Hyland, K. (2012) 'Bundles in academic discourse'. *Annual Review of Applied Linguistics* 32, 150–169

Ivanic, R. (1991) 'Nouns in search of a context: a study of nouns with both open-and closed system characteristics'. *International Review of Applied Linguistics in Language Teaching* 29 (2), 93–114

Jackson, L., Meyer, W., and Parkinson, J. (2006) 'A study of the writing tasks and reading assigned to undergraduate science students at a South African university'. *English for Specific Purposes* 25, 260-281

Johns, A. (2002) 'Destabilizing and enriching novice students' genre theories'. in *Genre in the Classroom: Multiple Perspectives*. ed. by Johns, A.M. Mahwah, N.J: Lawrence Erlbaum Associates, 237-246

Johns, A. (2011) 'The future of genre in L2 writing: Fundamental, but contested, instructional decisions'. *Journal of Second Language Writing* 20 (1), 56-68

Kachru, B. B. (1985) 'Institutionalized second language varieties'. in *The English Language Today* ed. by Greenbaum, S. Oxford: Pergamon Press, 211–226

Kachru, B.B., Kachru, Y., and Nelson, C. L. (2006) *The Handbook of World Englishes*. Malden, Mass, Oxford: Blackwell

Kalaskas, A.B. (2014) *Science lab report writing in postsecondary education: Mediating teaching and learning strategies between students and instructors*. Doctoral dissertation. Virginia: George Mason University

Kanoksilapatham, B. (2005) 'Rhetorical structure of biochemistry research articles'. *English for Specific Purposes* 24, 269-292

Kanoksilapatham, B. (2012) 'Research article structure of research article introductions in three engineering subdisciplines'. Professional Communication, *IEEE Transactions* 55 (4), 294-309

Kanoksilapatham, B. (2015) 'Distinguishing textual features characterizing structural variation in research articles across three engineering sub-discipline corpora'. *English for Specific Purposes* 37, 74–86

Lave, J. and Wenger, E. (1991) *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press

Street, B. and Lea, M. R. (2000) Student writing and staff feedback in higher education: an academic literacies approach. in *Student Writing in Higher Education: New Contexts* M. R. ed. by Lea, M.R. and Stierer, B. Buckingham: The Society for Research into Higher Education & Open University Press, 32-46

Leedham, M. E. (2011) *A corpus-driven study of features of Chinese students' undergraduate writing in UK universities* [online] PhD thesis. The Open University. available from < http://oro.open.ac.uk/29228/1297/Leedham_Thesis.pdf> [19 November 2019]

Leedham, M. and Fernández-Parra, M. (2017) 'Recounting and reflecting: The use of first person pronouns in Chinese, Greek and British students' assignments in engineering'. *Journal of English for Academic Purposes* 26, 66-77

Lillis, T. (2008) ‘Ethnography as method, methodology, and ‘Deep Theorizing’ closing the gap between text and context in academic writing research’. *Written communication* 25 (3), 353-388

Lim, J.M.H. (2019) ‘Explicit and implicit justifications of experimental procedures in language education: Pedagogical implications of studying expert writers’ communicative resources’. *Journal of English for Academic Purposes* 37, 34-51

Lin, L. and Evans, S. (2012) ‘Structural patterns in empirical research articles: A cross-disciplinary study’. *English for Specific Purposes* 31, 150–160

Lin, L. (2014) Variability in the rhetorical structure of research article introductions: The case of civil engineering. *Revista Española de Lingüística Aplicada* 27 (2), 405–432

Loudermilk, B. C, (2007) ‘Occluded academic genres: an analysis of the MBA thought essay’. *Journal of English for Academic Purposes* 6, 190–205

Mackey, A. and Gass, S.M. (2012) *Research Methods in Second Language Acquisition: A Practical Guide*. Oxford, UK: John Wiley & Sons

Marco, M.J.L. (2000) ‘Collocational frameworks in medical research papers: A genre-based study’. *English for Specific Purposes* 19 (1), 63-86

Martin, J. R. (1997) ‘Analysing genre: Functional parameters’. in *Genres and Institutions: Social Processes in the Workplace and School*. ed. by Christie, F. and Martin, J. R. London: Continuum

Martín, P. and Pérez, I.K.L. (2014) ‘Convincing peers of the value of one’s research: A genre analysis of rhetorical promotion in academic texts’. *English for Specific Purposes* 34, 1-13

Mathew, P., Nesi, H., and Vincent, B. (2019) ‘Corpus from scratch: Collecting and processing a sizeable EAP corpus in a (relatively) resource-poor context’. in *Proceedings of the 2017 BALEAP Conference: Addressing the state of the union: Working together = learning together*. ed. by Gillway, M. Reading: Garnet Education, 93-102

McGrath, L. (2016) ‘Open-access writing: An investigation into the online drafting and revision of a research article in pure mathematics’. *English for Specific Purposes* 43, 25-36

Melzer, D. (2009) ‘Writing assignments across the curriculum: a national study of college writing.’ *College Composition and Communication* 61 (2), 240-261

Merriam, S.B. and Tisdell, E.J. (2016) *Designing your study and selecting a sample. Qualitative research: A guide to design and implementation.* San Francisco, CA: John Wiley & Sons

Miller, C.R. (2000) ‘The Aristotelian topos: Hunting for novelty’. in *Rereading Aristotle’s Rhetoric*. ed. by Gross, A.G. and Walzer, A.E. Carbondale: Southern Illinois UP, 130-146

Montero, B. (1996) ‘Technical communication: Complex nominals used to express new concepts in scientific English-causes and ambiguity in meaning’. *the ESPecialist* 17 (1), 57-72

Moore, T. and J. Morton, J. (2005) ‘Dimensions of difference; a comparison of university writing and IELTS writing’. *Journal of English for Academic Purposes* 4 (1), 43-66

Moreno, A.I. and Swales, J. (2018) ‘Strengthening move analysis methodology towards bridging the function-form gap’. *English for Specific Purposes* 50, 40-63

Moskovitz, C. and Kellogg, D. (2005) ‘Primary science communication in the first-year writing course’. *College Composition and Communication* 307-334

Mudraya, O. (2006) ‘Engineering English: A lexical frequency instructional model’. *English for Specific Purposes* 25 (2), 235-256

Myers, G. (1985) ‘The social construction of two biologists’ proposals’. *Written Communication* 2, 219–245

Nathan, P. (2010) *A genre-based study of pedagogical business case reports*. PhD thesis. Birmingham: University of Birmingham

Nation, I. S. P. (2001) *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press

Nation, P. (2015) ‘Principles guiding vocabulary learning through extensive reading’. *Reading in a Foreign Language* 27 (1), 136-145

Nekrasova, T. M. (2009) ‘English L1 and L2 speakers' knowledge of lexical bundles’. *Language Learning* 59 (3), 647-686

Nesi, H. and Basturkmen, H. (2006) ‘Lexical bundles and discourse signalling in academic lectures’. *International Journal of Corpus Linguistics* 11 (3), 283–304

Nesi, H. (2014) 'Corpus query techniques for investigating citation in student assignments'. in *Corpus Analysis for Descriptive and Pedagogical Purposes.* ed. by M. Gotti, M. and Giannoni, D.S. Bern: Peter Lang

Nesi, H. and Gardner, S. (2006) ‘Variation in disciplinary culture: university tutors’ views on assessed writing tasks’. in *Language, Culture and Identity in Applied Linguistics.* ed. by Kiely,R., Clibbon,G., Rea-Dickins, P. and Woodfield, H. Sheffield: Equinox, 99-117

Nesi, H. and Gardner, S. (2012) *Genres across the disciplines: Student writing in higher education*. Cambridge: Cambridge University Press

Nesi, H. and Moreton, E. 'EFL/ESL writers and the use of shell nouns'. in *Academic Writing in a Second or Foreign Language*. ed. by Tang, R. London: Continuum, 126-145

Neumann, R., Parry, S., and Becher, T. (2002) ‘Teaching and learning in their disciplinary contexts: A conceptual analysis’. *Studies in Higher Education* 27 (4), 405-417

North, S. (2005) ‘Disciplinary variation in the use of theme in undergraduate essays’. *Applied Linguistics* 26, 431–52

Nini, A., Corradini, C., Guo, D., and Grieve, J. (2017) ‘The application of growth curve modeling for the analysis of diachronic corpora’. *Language Dynamics and Change* 7 (1), 102-125

Neuman, R., Parry, S., and Becher, T. (2002) ‘Teaching and learning in their disciplinary contexts: A conceptual analysis’. *Studies in Higher Education* 27 (4), 405–417

Ozturk, I. (2007) ‘The textual organization of research article introductions in applied linguistics: Variability within a single discipline’. *English for Specific Purposes* 26 (1), 25-38

Pan, F., Reppen, R., and Biber, D. (2016) ‘Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals’. *Journal of English for Academic Purposes* 21, 60-71

Parkinson, J. (2015) ‘Noun–noun collocations in learner writing’. *Journal of English for Academic Purposes* 20, 103-113

Parkinson, J. (2017) ‘The student laboratory report genre: A genre analysis’. *English for Specific Purposes* 45, 1-13

Parkinson, J. and Musgrave, J. (2014) ‘Development of noun phrase complexity in the writing of English for academic purposes students’. *Journal of English for Academic Purposes* 14, 48-59

Pawley, A. and Syder, F. (1983) ‘Two puzzles for linguistic theory’. in *Language and Communication*. ed. by Richards, J. and Schmidt, R. London: Longman, 191-226

Posteguillo, S. (1999) ‘The schematic structure of computer science research articles’. *English for Specific Purposes* 18 (2), 139–160

Randall, M. and Groom, N. (2009) ‘The BUiD Arab learner corpus: a resource for studying the acquisition of L2 english spelling’. in *Corpus Linguistics Conference (CL)*. held 20-23 July at the University of Liverpool. [online] available from < http://ucrel.lancs.ac.uk/publications/cl2009/> [16 July 2019]

Renouf, A. and Sinclair, J. (1991) ‘Collocational frameworks in English’. *in English Corpus Linguistics : Studies in Honour of Jan Svartvik*. ed. by Aijmer, K. and Altenberg, B. London: Longman, 128-143

Reppen, R. (2001) ‘Register variation in student and adult speech and writing’. in *Variation in English: Multi-Dimensional Studies*. ed. by Conrad, S and Biber, D. New York: Longman, 187–99

Rezoug, F. and Vincent, B. (2018) ‘Exploring lexical bundles in the Algerian Corpus of Engineering’. *Arab Journal of Applied Linguistics* 3 (1), 47-77

Rizzo, C.R. (2010) ‘Getting on with corpus compilation: from theory to practice’. *ESP World* 1 (9), 1-23

Römer, U. (2009) ‘The inseparability of lexis and grammar: Corpus linguistic perspectives’. *Annual Review of Cognitive Linguistics* 7 (1), 140-162

Römer, U. and O'Donnell, M. B. (2011) ‘From student hard drive to web corpus (part 1): the design, compilation and genre classification of the Michigan Corpus of Upper-level Student Papers (MICUSP)’. *Corpora* 6 (2), 159-177

Rosenfeld, M., Courtney, R., and Fowles, M. (2004) *Identifying the writing tasks important for academic success at the undergraduate and graduate levels GRE Board Research Report No. 00–04 R*. Princeton: Educational Testing Service (ETS)

Ruiz, L.P. (2006) ‘Unravelling noun strings: toward an approach to the description of complex noun phrases in technical writing’. *ES: Revista de filología inglesa* 27, 163-174

Samraj, B. (2002) ‘Introductions in research articles: variations across disciplines’. *English for Specific Purposes* 21, 1–17

Samraj, B. (2005) ‘An exploration of a genre set: Research article abstracts and introductions in two disciplines’. *English for Specific Purposes* 24 (2), 141-156

Schmidt, H. (2000) *English abstract nouns as conceptual shells: From corpus to cognition*. Berlin: Walter de Gruyter

Scott, M. and Tribble, C. (2006) *Textual Patterns*. Amsterdam: John Benjamins Publishing

Shi, L. and Dong, Y. (2015) ‘Graduate writing assignments across faculties in a Canadian university’. *Canadian Journal of Higher Education* 45 (4), 123-142

Simpson-Vlach, R. and Ellis, N. (2010) ‘An academic formulas list: New methods in phraseology research’. *Applied Linguistics* 31 (4), 487-512

Sinclair, J. M. (1991) *Corpus, Concordance, Collocation*. Oxford: Oxford University Press

Sinclair, J. M. (2004) ‘The search for units of meaning’. in *Trust the Text: Language, Corpus and Discourse.* ed. by Carter, R. London: Routledge, 19-33

Sinclair, J. M. (2005) ‘Corpus and text-basic principles’. in *Developing Linguistic Corpora: a Guide to Good Practice*. ed. by Wynne, M. Oxford: Oxbow Books, 1-16

Stoller, F.L. and Robinson, M.S. (2013) ‘Chemistry journal articles: An interdisciplinary approach to move analysis with pedagogical aims’. *English for Specific Purposes* 32, 45-57

Squires, G. (2005) ‘Art, science and the professions’. *Studies in Higher Education* 30, 127-136

Staples, S., Egbert, J., Biber, D., and Bethany, G. (2016) 'Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre'. *Written Communication* 33 (2), 149-183

Stubbs, M. (2002) 'Two quantitative methods of studying phraseology in English'. *International Journal of Corpus Linguistics* 7 (2), 215–244

Stubbs, M. (2007) 'Quantitative data on multiword sequences in English: the case of the word world' in *Text, Discourse and Corpora: Theory and Analysis*. ed. by Hoey, M., Mahlberg, Stubbs, M., and Teubert, W. London: Continuum, 163–89

Swales, J. (2019) 'The futures of EAP genre studies: A personal viewpoint'. *Journal of English for Academic Purposes* 38, 75-82

Swales, J. (2004) *Research genres: Exploration and Applications*. Cambridge: Cambridge University Press

Swales, J. (1996) 'Occluded genres in the academy: The case of the submission'. in *Academic Writing: Intercultural and Textual issues*. ed. by Ventola, E. and Mauranen, A. Amsterdam: John Benjamins, 45–58

Swales, J. (1990) *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press

Swales, J., and Feak, C. B. (2012) *Academic writing for graduate students: essential tasks and skills.* Ann Arbor, MI: University of Michigan Press

Swales, J. and Feak, C.B. (2004) *Academic writing for graduate students: Essential tasks and skills.* Ann Arbor, MI: University of Michigan Press

Swales, J. and Mustafa, H. (1984) *English for Specific Purposes in the Arab world.* Birmingham: University of Aston

Thomas, D. and To, V. (2016) 'Nominalisation in high scoring primary and secondary school persuasive texts'. *Australian Journal of Language and Literacy* 39 (2), 135-149

Tribble, C. and Wingate, U. (2013) 'From text to corpus: A genre-based approach to academic literacy instruction'. *System* 41, 307-321

Trimmer, J.F. (1985) 'Faculty development and the teaching of writing'. *WPA: Writing Program Administration* 2, 11-30

Varantola, K. (1984) *On Noun phrase Structures in Engineering English*. Turku: Turun Yliopisto

Vincent, B. and Nesi, H. (2018) 'The BAWE Quicklinks Project: A New DDL Resource for University Students' *Lidil* [online] 58. available from <http://journals.openedition.org/lidil/5306> [26 July 2018]

Ward, J. (2009) ‘A basic engineering word list for less proficient foundation engineering undergraduates’. *English for Specific Purposes* 28, 170–182

Wenger, E. (2015) *Communities of practice: A brief introduction* [online] available from <http://www.ewenger.com/theory/index.htm> [3 May 2019]

Winter, E. O. (1992) ‘The notion of unspecific versus specific as one way of analysing the information of a fund-raising letter’ in *Discourse description: diverse linguistic analyses of a fund-raising text*. ed. by Mann, W. and Thompson, S.A. Amsterdam: John Benjamins Publishing

Winter, E. O. (1977) ‘A clause relational approach to English texts: a study of some predictive lexical items in written discourse’. *Instructional Science* 6, 1–92

Wodak, R. and Meyer, M. (2010). ‘Critical discourse analysis: History, Agenda Theory and Methodology’. in *Methods of Critical Discourse Analysis*. ed. by Wodak, R. and Meyer, M. London: Sage

Wray, A. (2002) *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press

Wray, A. (2006) ‘Formulaic language’. in *Encyclopedia of Language and Linguistics*. ed. by Brown, K. Oxford: Elsevier, 590–597

Yang, R. Y. and Allison, D. (2003) ‘Research articles in applied linguistics: Moving from results to conclusions’. *English for Specific Purposes* 22 (4), 365-385

Zhu, W. (2004) ‘Writing in business courses: an analysis of assignment types, their characteristics, and required skills’. *English for Specific Purposes* 23 (2), 111-135

Appendix 1: List of Modules in the Civil Engineering and Quantity Surveying and Construction Management Programmes

Core Subject Modules in Bachelor of Engineering (Hons) in Civil Engineering

S. No.	Semester of Study	Module Code	Module Title	Corpus Holdings	Genre
1	1	CIVL 0001	Introduction to Civil Engineering	4 (shared across the two programmes)	Lab Report Case Study
2	2	CIVL 0002	Engineering Surveying 1	3 (shared across the two programmes)	Lab Report
3	2	CIVL 0003	Civil Engineering Construction 1	3 (shared across the two programmesc ommon)	Case Study
4	2	ENGG 0002	Design and Visualisation	0	Only diagrams (no writing at all)
5	2	CIVL 0004	Materials 1	2	Case Study Lab Report
6	3	CIVL 0005	Applied Geology	7	Lab Report Site Investigation Report
7	3	CIVL 0006	Structural Mechanics 1	0	***
8	4	ENVT 0002	Introduction to Environmental Engineering	6	Explanation Case Study
9	4	CIVL 0007	Civil Engineering Practice	5 (shared across the two programmes)	Manual Case Study
10	4	CIVL 0008	Hydraulics 1	18	Lab Report
11	4	ENGG 0003-CIVL	Project 1	0	***
12	5	ENGG 0004	Project Management	15 (shared across the	Exercise

				two programmes)	
13	5	CIVL 0009	Geotechnology 1	5	Lab Report Exercise
14	5	CIVL 0010	Materials 2	1	Lab Report
15	6	CIVL 0011	Highway and Transport Engineering	2	Case Study
16	6	CIVL 0012	Structural Design	0	***
17	6	CIVL 0013	Structural Mechanics 2	1	Exercise
18	7	CIVL 0014	Railway Engineering	3	Case Study
19	7	CIVL 0015	Structural Mechanics 3	7	Exercise
20	7	CIVL 0016	Geotechnology 2	4	Exercise Lab Report
21	7	ENGG 0005-CIVL	Project 2-Part A	0	***
22	8	CIVL 0017	Civil Engineering Design	0	***
23	8	CIVL 0018	Hydraulics 2	0	***
24	8	CIVL 0019	Concrete and Concrete Structures	0	***
25	8	ENGG 0006-CIVL	Project 2-Part B	0	***

Core Subject Modules in Bachelor of Science (Honours) in Quantity Surveying and Construction Management

S. No.	Semester of Study	Module Code	Module Title	Corpus Holdings	Genres
1	1	CIVL 0001	Introduction to Civil Engineering	4 (shared across the two programmes)	Lab Report Case Study
2	2	CIVL 0002	Engineering Surveying 1	3 (shared across the two programmes)	Lab Report

3	2	CIVL 0020	Construction Technology and Services 1	1	Explanation
4	2	ENGG 0002	Design and Visualisation	0 (shared across the two programmes)	Only diagrams (no writing at all)
5	2	CIVL 0021	Building Strength and Stability-1	1	Explanation
6	3	CIVL 0022	Economic Background to Construction	1	Explanation
7	3	CIVL 0023	Internal Environment	1	Explanation
8	3	CIVL 0024	Domestic Design and Construction	3	Explanation
9	4	CIVL 0007	Civil Engineering Practice	5 (shared across the two programmes)	Manual Case Study
10	4	CIVL 0025	Quantity Surveying Practice - Building	0	***
11	4	ENGG 0003	QSCM - Project I	0	***
12	5	ENGG 0004	Project Management	15(shared across the two programmes)	Exercise
13	5	CIVL 0026	Contractual Procedures and Estimating	3	Case Study Explanation
14	5	CIVL 0027	Building Strength and Stability 2	17	Lab Report
15	6	CIVL 0028	Property Surveys and Building Pathology	1	Case Study
16	6	CIVL 0029	Cost and Quantitative Management	6	Case Study Explanation
17	6	CIVL 0030	Quantity Surveying Practice (Civil)	6	Case Study Explanation
18	7	CIVL 0031	Innovation in Construction	3	Case Study
19	7	CIVL 0032	Commercial Management	3	Case Study
20	7	CIVL 0033	Contract Management R	3	Explanation
21	7	ENGG 0005	QSCM - PROJECT 2-Part A **	0	***
22	8	CIVL 0034	Construction Finance	0	***
23	8	CIVL 0035	Procurement Pathways	0	***
24	8	CIVL 0036	Management Systems in Construction	0	***
25	8	ENGG 0006	QSCM - PROJECT 2-Part B	0	***

Appendix 2A: Move structure of Methodology Recount assignments in OCAW-CE

Moves and Steps	Examples	Percentage of assignments
Introduction Move 1 (Move I1)		
Step 1	**Stating the importance of the topic** • *It is useful for locating drains, rods, dams and other earth structures.* • *The shear strength is one of the most important engineering properties of a soil, because it is required whenever a structure is dependent on the soil's shearing resistance.*	33.3
Step 2	**Referring to established knowledge** • *The phenomenon of capillary effect can be explained microscopically by considering cohesive forces and adhesive forces.* • *Finding the density of a metal sample can help to determine its purity.*	73
Step 3	**Providing diagram**	22
Step 4	**Stating aim** • *The **objective** of this lab experiments is to find out the tensile strength of the given steel specimen.* • *The **purpose** of this experiment was to determine the dynamic viscosity of a given fluid*	100
Move 2 (Move I2)		
Step 1	**Stating the method**	57

	• *Testing proceeds by displacing the lower half of the split box and measuring the horizontal shear forced (T) transmit through the soil to the upper portion of the box.* • *The theoretical discharge is calculated using the standard equation derived from Bernoulli's theorem.*	
Step 2	**Justifying method** • *Soil particle size distributions test is faster than other test to determine the engineering properties of permeability, compressibility and shear strength*	3
Methods		
Move 1 (Move M1)		
Step 1	**Listing apparatus/equipment** • *Different type of sieves, balance: to measure the soil sample weight, sieve shaker, cleaning brush and mixer or blender.* *a) Measuring beaker* *b) Eureka can* *c) Density bottle*	93
Step 2	**Recounting experimental procedure** • *Wipe every example to a surface-dry condition, and uproot any free sand grains or incrustations from the faces that will be in contact with the bearing pieces of the testing machine.* • *Write down the weight of dry soil in balance (about 500 grams).*	93
Results		
Announcing Results (Move R1)		
Step 1	**Using location statements** • *The above graph shows the relation between the sieve size and the percentage of sample passing through each sieve size.*	49.2

	• *As seen from table 1, wooden blocks can take loads approximately 9 times when the load is in a perpendicular direction to its fibres.*	
Step 2	**Displaying results in a figure/table**	88.8
Step 3	**Detailing data analysis process** • *…calculate the compressive strength as follows: Compressive strength= P/A* *The coefficient of permeability is standardized at 20°C, and the permeability at any temperature T is related to K20.*	19.04
Step 4	**Calculating results** • *Bulking factor of sand= [Maximum increased volume - 1] = [465/325 – 1] = 0.430*	87.3
Step 5	**Reporting results** • *The results show that the viscosity for glycerin is 1 x 10-3, for automobile oil is 3.7 x 10-4, and for cooking oil is 1.1 x 10-4…* • *Result: Dry density=17.663 g/ml.*	73
Discussion		
Contextualising Discussion (Move D1)		
Step 1	• *The Point at which an increase in strain occurs without an increase in the stress is defined as the yield point.* • *The viscosity of relatively high viscosity fluids, such as oil, glycerin, and castor oil can be determined.*	38
Interpreting results (Move D2)		
Step 1	**Stating selected findings** • *From the overall results of analysis, the Bulk density of sand is 13.765 KN/ m3 and Dry density of soil13.538 KN/ m3.* • *For this experiment KT equal to K_20 = 0.03 cm/sec.*	66.6
Step 2	**Interpreting results**	46

	• *There is some variation in values due to surface tension of water.* • *…which mean that glycerin is more viscous than the two oil, because of this the ball took long time when it fall in the glycerin.*	
Step 3	**Accounting for unexpected outcomes** **Stating sources of error** • *Possible source of error: Slipping of the steel at the testing machine grips*	11.1
Conclusion		
Summarising the study (Move C1)		
Step 1	**Stating selected findings** • *…the total mass sample after sieving is 497g is less than total mass 500g before sieving where 3 g of soil is missing.*	66.6
Step 2	**Drawing Conclusions** • *The results of viscosity and specific gravity are within the acceptable values.* • *The Soil classification as per the British Standard (MIT soil classification) is Sandy Gravel (G S).*	77.7

Appendix 2B: Move structure of Case Study assignments in OCAW-CE

Moves and Steps	Examples	Percentage of assignments
Introduction		
Move I1 Introducing the topic		
Step 1	**Providing background information** • *Gulf Company Council (GCC) is a regional Intergovernmental political and Economic union consisting of all Arab states of the Persian Gulf, with the exception of Iraq.* • *The process of the transport planning will go through different phases:*	96.7
Step 2	**Giving definitions** • *Sustainability is 'the open and unrestricted and conception of sustainable development' (Jacobs, 1999:33).* • *Corrosion means atmospheric oxidation of metals.*	58
Step 3	**Claiming importance** • *In order to cope with rapid urbanization, meet with the requisites of burgeoning trade, encourage sustainable development, and to diversify the region away from reliance on oil and gas in an attempt to boost industrial development…* • *Planning the transportation is very crucial in order to achieve sustainable transportation system.*	87
Step 4	**Stating the broad problem** • *Therefore, sustainable construction is considered as one of the main challenges of sustainable development (Hoffman, 2008).*	83.8

	• *This is leading to a high consumption of natural resources, increasing the emission of greenhouse gases and polluting the ecological system...*	
Move I2 Stating the aim of the study		
Step 1	• *The main **objective** of this assignment is to write about mega project in Oman or in other country in the world.* • *This report **aims** to identify the principles of sustainability and the main factors that create sustainability in buildings.*	96.7
Move I3 Outline of the assignment		
	• *This paper has progressed in a sequence of phases: identifying the principles of sustainability through conducting a literature review; mapping out the important factors that create elements of sustainability in buildings in Oman according to the codes and standards; description of the selected case study which is chosen as sustainable project; analytical comparisons between the model criteria and the case study.*	80
Introducing the case study		
Move CD1 Description of the case		
Step 1	**Explanations with/out citations** • *…some materials during their life cycle produces emissions that harms the environment and increases the temperature of the earth.* • *The weight of the soil used for reclamation causes the water in the clay beneath to move outward along the sand piles.*	100
Step 2	**Description of the case**	100

	• *The road is starting from petrol station and ending by the roundabout that leads to expressway.* • *The Metro Project comprises of six (6) Lines are first line starting from north of King Salman Bin Abdul Aziz Street and ending at Dar Al Badia neighborhood in the south and runs in the North-South direction along Olaya and Bath streets.*	
Step 3	**Providing numerical data** • *The estimated cost is $14billion and the total cost (budget) was $20billion.* • *The R165S has a 50-hp Perkins diesel turbocharged…* • *In 2006, thirty seven million cubic meters were being dealt with and used again.*	100
Step 4	**Displaying descriptions in tables/graphs/figures**	83.87
Move CD2 Indicating the problem/evaluation to be conducted		
Step 1	• *This road experienced many secondary levels of accidents like Rear-end collision and Sideswipe accidents.* • *According to the factors identified before, performance of this greenhouse is next analyzed to test the sustainability concepts on the building according to the international standers and codes that have been identified in tables 1 to 12.*	100
Results		
Move R1 Announcing the findings		
Step 1	**Reporting findings with/out reference to the literature/benchmarks** • *The table 3 shows the envelope of visibility and different design speed for full overtaking sight distances. From the results we have obtained that design speed for selected road is 85A.*	100

	• *…the building's primary source of energy is the Solar Panels.*	
Step 2	**Displaying findings in tables/graphs/figures**	87
Step 3	**Interpreting the results obtained** • *Green-Nest is performing well in this criteria, as it has light fixtures, fittings and luminaires.* • *This satisfies many standards such as National Association of Home Builders (2014).*	83.8
Conclusion		
Move C1 Concluding the study		
Step 1	**Making recommendations** • *Therefore, the construction company that has an innovative idea should send its laborers to take courses in other countries know about this area.* • *Therefore, there are many suggested solutions to improve the traffic process such as other types of intersection (single-lane dualling), traffic lights or roundabout to the busiest exits (SQU & Knowledge Oasis).*	96.7
Step 2	Restating the purpose/summarising the study • *This project discussed the innovation of construction in three buildings in Oman including…* • *In this report I explain more information about the Lupu bridge.*	100

Appendix 3A: Sample Assignment Brief (excerpt) of a Lab Report

Department of Civil Engineering

Introduction to Civil Engineering -CIVIL 0001

Individual Assignment (Case Study) **Date of submission -27th December 2015**

OBJECTIVE:

The fundamental objective of this assignment is to demonstrate the student ability to

1. Identify, and analyse the properties of building material
2. Discuss structural behaviour of the materials on the building components..

OUTCOMES:

1. Discuss the characteristics, properties and use of materials (Learning outcome no.2 MIG CIVL0001).
2. Explain the effects of structural behaviour on construction components (learning outcome no.3 MIG CIVL0001).

TASK:

Select one of the following projects.

1. Collect a sand sample (preferably as in site) and conduct
    - Bulk density
    - Silt content
    - Sieve analysis

    Discuss the different properties of your sample based on your result analysis.

2. Get at least two different types of cement and conduct the following tests to identify the properties of your sample.

- The Consistency test; Fineness of cement, initial and the finial setting time tests

3. Collect two samples of coarse aggregate and discuss their properties based on the shape (elongation and flakiness index). How these properties effect on the workability of the mix of concrete of 1: 2: 4 and 0.7 W/C ratios using slump cone test
4. Do slump and compaction factor tests using any type of aggregate with different W/C ratios 0.6, 0.7, and 0.8, and discuss your results, and how the W/C ratio affect the concrete properties.

Note: These are some examples where students still can propose different topics for their case studies after module leader approval.

Please keep in mind following while writing your report:

- The report is often used to present, analyze and summarize your data.
- Follow the following instructions on writing your report:

Report Format

- The size limit: the number of pages should be from 4 to 10 pages, additional pages can be used as appendix.
- Generally use 1.5 x line spacing unless advised.
- Start with the title page including student name and ID number as requested.
- Include your own field sketches or photographs. It is generally better to place figures within the text rather than group them at the end of the report.
- Give a sequential number to all figures (Fig. 1, Fig. 2, Fig. 3, etc.)
- Add a brief, informative caption to the figures.
- Give a sequential number to all tables (Table 1, Table 2, Table 3, etc.), with informative title

Report Sections

The report should contain the following sections:

- Abstract
- Introduction

- The materials used
- Structural behaviour
- Conclusions
- References

1. **Abstract (maximum 1/2 page)**

The abstract must be brief, but it has to tell the reader

- What is the goal of the work (e.g. The objective of this study is to study the influence of w/c ratio on concrete workability) etc.
- Which type of work has been carried out (e.g. **lab experiments** and methods)?
- What are the results obtained (e.g. increasing the w/c ratio increased the concrete workability), preferably not more than 2 sentences.

2. **Introduction (maximum 2 page)**

- Discuss in general the material classification, and the criteria of selecting the materials.
- The material properties (mechanical, physical, production and construction, and the aesthetic, etc.)
- The objective of the study.
- Give the general plan of the report (e.g. how do you collect your sample and what will you do to discuss these properties).

3. **The material (your material) (maximum 5 page)**

- Discuss the properties of the selected material (e.g. for aggregate: the practical shape, surface texture, toughness, abrasion....etc. for cement: production, hydration, fineness,.. etc.)
- The tests to check the properties.
- The results you get and analysis for those results.

4. **The structural behaviour (maximum 2 page)**

- Explain how the material will affect in terms of behaviour in beams, columns, structural frames, pads and machine bases, floors, etc.

5. **Conclusions (maximum 1/2 page)**

State briefly what are the major achievements of your work.

6. References

List alphabetically your references which should include: Journal papers, proceedings, books, book chapters, internet sources etc.

7. Appendix

Any additional figures, tables, and photos.

DELIVERABLE:

1. The report should not exceeding 10 pages.

GUIDELINES:

Follow the guidelines mentioned below for your assignment.

- Assignment should have a Title Page. Title Page should contain the following information:
  - Assignment Name
  - Class
  - Student name
  - Student ID
- Use page numbers
- Assignment should be typed in your own words and computer typed using Times New Roman font size 12.
- Heading should be with Font Size 14, Bold, Underline
- Use Diagrams and Examples to explain your topic.
- Copy and paste from the Internet sources is strictly not acceptable and will be treated as a plagiarism and will be dealt with as per polices given below.

Appendix 3B: Sample Assignment Brief (excerpt) of a Case Study

Department of Civil Engineering
Civil Engineering Practice
CIVL 0007
Group Project
Date of submission - 12/12/ 2015

OBJECTIVE:

•Providing a thorough understanding of Requirements and important factors in an engineer's role to create and maintain sustainable environment via a real-life scenario project.

OUTCOMES:

One of the most significant roles of an engineer is to create a sustainable environment. In order to fulfill his/her role, an engineer must act according to the principles introduced by ICE and Royal Academy of Engineering. Through investigating the real applications of sustainability in engineering project, students will be able to compare the principles in theory and practice. This knowledge will enable students to obtain a profound understanding of sustainability in engineering work which leads to applying this knowledge in their career and train them to become involved as engineer in sustainable and environment friendly construction.
Hence the outcome of this project is for the students to understand the engineer's role in creating and maintaining a sustainable environment.

The purpose of this coursework is to understand to what extent local sustainable projects in Oman are actually performing up to the standards defined in the scientific literature with regards to sustainable development in construction and built environment. To do so literature review must be conducted to be able to first understand the principles of sustainability. Moreover a model should be prepared to define appropriate factors and criteria in order to be able to measure the performance of the case-study.

Students must complete the steps in this task as following:
1. Students are required to first study the principles of sustainability and sustainable development in construction and built environment introduced in the literature and references. Concepts of sustainability, sustainable development must be reviewed, discussed, via exploring academic and technical journals, technical reports, papers, conference proceedings, text books, case studies, government and professional guidelines in the field of construction and sustainability.

2. Furthermore through reviewing this literature a model with appropriate factors and criteria should be prepared that defines sustainability in construction and later can be used to compare and measure the case study to understand to what degree the case study is according to this model's criteria.

3. Next step is to identify a civil engineering project which is a candidate for case-study. This project must be appropriate in the sense that it can be studied by students in order to find the implemented elements of suitability. This case-study can be construction project that is recognized to have implemented sustainability concepts.

4. Students must introduce the case-study and identify what activities have taken place in that specific project, to provide sustainability in terms of environmental, social, cultural, financial, etc. This project maybe in any areas of civil engineering work the data can be collected through research i.e. via contacting consultant and contracting companies, literature, interviews (structured/semi-structured, questioners), surveys, etc.

5. Next step is to critically examine the case-study using the principles, model factors and criteria identified in the first step. You must analyze the performance of the case-study and measure to what degree it is up to the standards you defined in your model with respect to sustainable construction.

6. Next step is to prepare a report. Select appropriate headings and subheading to guide the reader. Remember that the reader will only understand as much from your work as you guide them. Use your writing skills to attract the reader. This report must be in the following format and not exceed 5000 words.

To summarize the report includes the following steps:
1. Literature review; to define the criteria required to measure performance of the case-study
2. Identifying the case-study , introducing the case-study and find relevant information
3. Critical analyze of the performance of case study

Important notification
• You must have citations from reputable journals, articles, books, etc. in your paper. this is considered of great importance in the quality of your report and hence your mark
• Note that mark is given to quality of writing and just by completing the writing you will not be entitled to full mark of that part
• In case your report covers the relevant areas with better quality the possibility of receiving a higher mark increases
• Avoid using irrelevant and general websites as references.
• Your assignment is a report therefore high degree of your mark depends on the quality of language of your writing. Avoid using Google translator.
• You can use the support of academic writing unit in language center to enhance your report in grammar, vocabulary; referencing, etc. also note that they conduct workshops from time to time in these subjects.
• Pay great attention to the referencing, MEC preferred method of referencing is CU Harvard.

Coursework Proposal

You must prepare and submit your CW proposal through the Moodle the week after the assignment is given to you. This proposal must contain the following:
Title of the Assignment, Module Name and Title and Semester (Fall 2015)
Information regarding your group members
Explanation of your understanding about the project and tasks you shall carry out Task allocation among the group members
Project program (time schedule)

See the FORMAT of the proposal in the Assignment
Important note:
• Students must note that Marks for the assignment will depend only on the presentation (viva). Viva is compulsory and NO marks will be awarded to the assignment if student does not appear in the viva on the specified date. This means that in the presentation day if the student fails to explain properly about his/her role in the work, he/she will get zero mark.
• Moreover students must submit a coursework proposal to show their understanding about deliverables/ time lines, etc.
• To assess individual work of each student a viva factor mark will be given to student's performance in viva session. Students must demonstrate adequate knowledge and explain accurately about all parts of project and particularly about their role in the project.
• **Note that final report mark is will be adjusted using Viva Factor to incorporate every student's individual contribution to the group work**

Proposal Format
Follow the guidelines mentioned below during writing your report:
☐ The report should have a Title Page. Title Page should contain the following information.
☐ Title
☐ Class
☐ Student name
☐ Student ID
☐ Use page numbers
☐ Use your own words on writing, Times New Roman font size 12
☐ Heading should be with Font Size 14, Bold, Underline

Report Format
• Number of pages: Maximum 20 pages excluding references list and the appendix (if any).
• Line spacing: Generally use 1.5 x line spacing.
• Start with the title page (including your name) or student number as requested.
• It is preferable to place figures within the text rather than group them at the end of the report.
• Give a sequential number to all figures (Fig. 1, Fig. 2, Fig. 3, etc.)
• Add a brief, informative caption to the figures.

The Report should contain the following sections:
• Table of contents
• Introduction
• Report body
• References

1. Table of content

Include heading, subheading, and page numbers.
2. Introduction

The introduction is very important as it sets the context for the report. Summarize the brief, briefly outline the case and focus on its significance (problem statement), state the work's aims and objective, and describe how the report is organized.
3. report body
This section will contain the review of literature, explaining the case study, method of research, data obtained and discuss the finding of the case study. Headings should be informative and descriptive providing a clue to the contents of the section.
4. Conclusions (maximum 1/2 page)

Every report should include a concluding statement/s on the subject of the report. Restate the aim of the report and state how you have achieved it. Present the main findings and key recommendations in a summarized. You should also restate the limitations of the report.
5. References

This is a list of all the sources of information you have referred to in the report.
6. Appendix

Any additional figures, tables, and photos if required
GUIDELINES:
Follow the guidelines mentioned below during writing your report:
The report should have a Title Page. Title Page should contain the following information.
Title
Class
Student name
Student ID
Use page numbers
Use your own words on writing, Times New Roman font size 12
Heading should be with Font Size 14, Bold, Underline
Use Diagrams and Examples to explain your topic.
Copy and paste from the Internet sources is strictly not acceptable and will be treated as a plagiarism and will be dealt with as per polices given below.

Rules & Regulations
• If any content or diagram of the report body is found to be copied from the other students then marks will be deducted from both reports.
• The purpose of the report is to do some research work so you can consult books in Library or use internet or computer magazines or any other source.

Appendix 4: Sample Interview Transcript (Lecturer No. 4 from the Civil Engineering Department, MEC)

Interviewer: Good Morning Dr X. As discussed a bit earlier I will be asking you a few questions on assignment writing in your department.

Interviewee 4: Good morning

Interviewer: How important is assignment writing in your department?

Interviewee 4: Very important…it covers 50% of the total assessment

Interviewer: What levels of modules do you teach?

Interviewee 4: I've taught Level 1 but currently I'm teaching Levels 2 and 3.

Interviewer: Could you tell us about the different types of written assignments that you design for your students like from your assignments I've seen that you have lab experiments. What other kinds of assignments do you have?

Interviewee 4: So, the modules normally I…Geotechnology 1…Geotechnology 2 more into soil properties so there are lab reports for Geotechnology 1 and Design Reports for Geotechnology 2, so in the Lab Reports students are…they are asked to have their own soil samples, conduct some soil tests in the lab, then discuss and they also need to provide conclusions based on their own judgement regarding the results. So in Geotechnology 2 I just give them design tasks design this…design retaining walls …design shallow foundation…deep foundation…for those things they need soil data …they need to go for getting soil sample, do the lab-testing in the lab and get the results…use those results for the design but in Geotechnology 2 we don't assess their lab…we just assess their design report projects.

Interviewer: So there's a lot of writing involved there?

Interviewee 4: A lot of writing, a lot of writing

Interviewer: and also a lot of calculations?

Interviewee 4: A lot. The issue is normally we limit the data they get because they have to generate other datas from the given data…the other issue is they have to use their common sense …their own judgment in their discussions…even if they get some unrealistic data they need to discuss …realistic or unrealistic…why…from previous studies or from some books…

Interviewer: So they have to do some kind of literature review also in the design report?

Interviewee 4: Yes because you cannot get all the data from the design …so from the given data there are derivations to generate the missing data …for that they need to refer books…and they need to cite and they need to justify their results

Interviewer: So they have a separate methodology chapter also?

Interviewee 4: Yes, normally we give them the format…how they need to write the conclusion, results, discussion, interpretation, those things.

Interviewer: So what type of report is this?

Interviewee 4: We call it Design Report. It has all the components of report writing.

Interviewer: So, what is the most frequently used type of assignment in your field…here in MEC

Interviewee 4: In Civil, especially in Level 2 and Level 3 they need to provide design reports. Okay, in Level 1 they need to provide some Site Investigation reports, Case Study Reports but as they go up…there are lots of design reports. Normally we ask them to produce design reports

Interviewer: At the higher levels? As I finished collecting your assignments for one semester, I found that there are lots of assignments in Level 2 as compared to the other levels. Why is it that there are more modules in Level 2 as compared to the other levels?

Interviewee 4: Yeah, in Level 1 we ask students to produce lab reports. That is not there in Level 2 because they have done it in Level 1. So Level 3 students they don't have to repeat it because they have already been assessed for that part of the component.

Interviewer: But in Level 2 there are quite a few…

Interviewee 4: Assessment reports…that's why. In Level 2 we have two assessment components- the lab reports and the design reports. In Level 3 we only have the design reports

Interviewer: Do you set particular type of assignments depending on the level of the students?

Interviewee 4: Exactly.

Interviewer: Oh, you just answered this question. Do you have group and individual assignments?

Interviewee 4: Only individual assignments

Interviewer: Especially in the modules you teach?

Interviewee 4: Yes

Interviewer: Suppose I were to ask you. A case study and a site investigation report. How do they differ?

Interviewee 4: For example for the site investigation, they have to do actual site tests …they need to go to the site, take some equipment tools and then they need to drive into the soil and get the soil data. In case studies we just ask them either we give them the data or gather the already existing data

Interviewer: And they have to make some recommendations regarding the case?

Interviewee 4: Yes

Interviewer: What kind of difference do you expect between the writing of Level 1 or Level 0 students and the writing of Level 3 students on the other hand?

Interviewee 4: I think it improves as you go up as they are getting more experience because in each module we push them to follow the standard way of writing…we give them formats…they need to

discuss…they need to interpret the results …they need to conclude…so those things as the level goes up it gets improved.

Interviewer: So what do you value the most in student work?

Interviewee 4: Normally in my modules I give more attention to the analysis and result interpretation part. If the analysis is ok and based on the results they interpret those things, I give more marks to those parts of the components

Interviewer: And do you think language is important?

Interviewee 4: Very important. But the problem…the grammar…for example even the words they use…the issue here is they Google because English is not their first language they…

Interviewer: You mean they use the Google Translator tool?

Interviewee 4: Translator tool. With the translator tool they don't sometimes they get some strange words which is not appropriate for that particular…

Interviewer: Can you give me an example?

Interviewee 4: For example, I teach them stresses in soil mass. Stress…they give me emotional stress and psychological stress which has nothing to do with soil stress (laughs)

Interviewer: So this is a real problem because they don't know how to use the tool appropriately. So these are some of the things that you dislike finding in students' work? Firstly, the language and what else you dislike…

Interviewee 4: Normally the other thing is they do the analysis they get results…they put some results that is unrealistic. So they need to use their common sense whether that value is right value or not. I usually push them. Get something…check…use your common sense whether that result is right or not and that thing…for example, the unit weight of soil…they give a very strange…normally it has to be 10K per Newton cube or 20 kilo per Newton cube. Sometimes they give 30, 000 or 40,000…so that one I don't like

Interviewer: So you think your assignments are designed in such a way that they effectively assess students on the qualities you value?

Interviewee 4: Yes

Interviewer: Do you have any particular ways and means of helping students with the written part of their work?

Interviewee 4: So normally in this report writing normally we give them the formats …the way the methodology should be written ...the way the discussion should be written…the conclusion …they need to relate their objectives to the results they get in their conclusions. We usually advise them write this way use these words avoid these personal pronouns those things …they come to me I just advise them in this way but the grammar most of the time is a big issue

Interviewer: So who do you think can support their writing problems like is it the language teachers or is it the subject teacher like yourself?

Interviewee 4: I think both because at the language level there should be some tools to assess the students' language ability. When they come to the subject matter part because there are some technical jargons so the issue I raise they use emotional stress psychological stress. Those things should be avoided at the Civil level. CE level. The English proficiency level has to be …

Interviewer: Before they enter you mean?

Interviewee 4: Yes, there should be some screening mechanism.

Interviewer: Now let us suppose that it's not happening. So what is the solution we have now for this…because it may not be possible…the Foundation Programme duration is one year and students come in with a 2 or a 2.5 IELTS and it is not possible for them to progress to a 5.5 within the duration of a year. So these are the kind of students we have…so, what solution can you think of?

Interviewee 4: Naturally, if you have some commitment to change, you can change at any level. I think we need to bring some attitudinal change within our student cohort so this is an English medium institution so English should be given more weightage. So every student should improve their English proficiency so at this level we need to push them…we need to encourage them…we need to persuade them so that the current English using is improved. After they pass the Foundation, at each level encouragement should be there at each level. Level 1, level 2, Level 3

Interviewer: There should be continuous support

Interviewee 4: Yes, continuous support. The issue here is the marks we give them …you know why this attitudinal change? Students who score 40 plus out of 50 …they don't want to go for A grade ..they want to pass…normally they should have tried to score good grades in the final also so…that thing is not there…it has to be improved…so maybe some attitudinal should be put in place in my view. Student should go for something better…why be satisfied…

Interviewer: Do you think students are clear about assignment expectations?

Interviewee 4: Yes, normally we put the marks…mark distributions also so I think…

Interviewer: Do they understand what they need to do in their assignments?

Interviewee 4: Sometimes they have that problem but they come to us …we discuss in the class also. They ask 'What do you expect me to do in Question Number One? We discuss in the class. 'You are expected to do this…do that'…So when they have some problems they come…normally they come

Interviewer: So let's talk about some strategies that students use to complete their written work. Do you think they borrow assignments from previous semesters as models or maybe to plagiarize? What do they do with those?

Interviewee 4: Sometimes…it is very rare…they just copy from the previous reports…but they give you wrong data. We can identify it. They get a penalty for it. They need to resubmit.

Interviewer: So when students plagiarize, this is what happens? So, now we have language teachers and subject teachers, can you think of a solution? I'm repeating myself I know…What kind of

solution can we come up with to provide this continuous writing support and who should be involved in that support? Is the language teacher or …

Interviewee 4: When they are in the language level…Foundation…they don't know what is coming …we have the Mathematics issue also. Some of my modules, Mathematics is the pre-requisite …so, normally we talk to those mathematics people and …'Give more attention to these areas because these areas are required for these modules …some kind of contact is there

Interviewer: But with the language, do you have that kind of contact?

Interviewee 4: No

Interviewer: Not at all…Why is that? There is no collaboration you feel?

Interviewee 4: Yeah, I'm contacting the Mathematics because Mathematics 2 is the pre-requisite for Geotechnology 2…clearly written there …so I've to talk to them.

Interviewer: Let me just come back to your coursework in Geotechnology 2 where there are lots of questions given and it has an Introduction and it has a Conclusion, so what do you expect students to do in that assignment?

Interviewee 4: For example, those questions are related to design and structure. In that case they need to introduce the coursework. In this coursework these these these things are to be covered and these activities are to be conducted. Finally they should go to the lab…then they should put the lab results …the soil data results should be…

Interviewer: And what should they conclude with? I've seen a lot of writing in the conclusion section.

Interviewee 4: Sometimes even the conclusion is very discursive and outside the…it is not even related to the topics

Interviewer: But what do you expect them to do in the conclusion?

Interviewee 4: In the Conclusion I expect them to…there are the tasks that are given to them. 'In this task we did this we did this and some things are wrong because data were wrong or when we measured the things equipment was defective …

Interviewer: some reflection

Interviewee 4: Yes, reflection because you know normally we cannot get things perfectly …so what are the limitations in their work …where did they go wrong? They should specify that…

Interviewer: thank you very much. It was very informative.

Interviewee 4: Thank you also.